

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

**THE SEARCH FOR A
ROUTINE OUTCOME MEASURE FOR
MULTIDISCIPLINARY INTERVENTIONS
WITH YOUNG PEOPLE
IN NEW ZEALAND**

A thesis presented in partial fulfilment of the requirements for the degree of
Doctor of Clinical Psychology
at Massey University, Manawatu campus, New Zealand.

Amber Barry

2011

ABSTRACT

Routine outcome measurement in mental health services is becoming standard practice around the world and has benefits for the public, researchers, funders, managers, and most importantly clinicians and clients. New Zealand's child and adolescent mental health services mandated the use of the *Health of the Nation Outcome Scale for Children and Adolescents* in 2005. Unfortunately, this decision was not based on a systematic or comprehensive review of available measures, as had been the case overseas. This research aimed to identify and recommend a measure that was psychometrically robust, feasible, and useful to multidisciplinary practitioners in New Zealand. The six studies conducted also contributed to the scarce literature on practitioner's opinions about routine measurement and the suitability of an overseas measure for a New Zealand population. A postal survey and focus groups with practitioners indicated mixed support for outcome measurement, with few measures identified or recommended by participants. Although many barriers were endorsed, the reasons to use measures were generally rated as more important. An extensive search of the international literature provided an up-to-date review of outcome measures for young people. From the initial 1665, six short-listed measures were reviewed by experts and practitioners before the *Ohio Youth Problems, Functioning, and Satisfaction Scales* was recommended as most suited for multidisciplinary interventions. Two further studies provided preliminary data about the Ohio Scales' (developed in the USA) suitability for use in New Zealand. Consultation with clinicians, parents, and young people indicated minor wording changes would improve its appropriateness for New Zealand. Then, a New Zealand field-test indicated the psychometric properties of the measure were preserved, but significantly different scale scores were found in comparison to a community sample from Ohio, USA, although this could have been a product of the small sample size. The Ohio Scales offers clinicians the necessary balance between breadth of content, brevity, and psychometric strength for its use to benefit their practice and the organisations they work in.

ACKNOWLEDGMENTS

This thesis has involved many people and I am grateful for their assistance, support, and guidance. Bringing together the variety of studies that make up this research could not have happened without the input of university staff, research and clinical colleagues, family, friends, and, most importantly, the organisations and individuals who took part. I would like to acknowledge their input here.

Thank you to my thesis supervisors, Prof. Ian Evans, Dr Shane Harvey, and Ms Cheryl Woolley, and the university staff who assisted me through this process. I am grateful for the financial support that made this research possible. This included funding from the High and Complex Needs Unit, the School of Psychology (Graduate Assistantship and doctoral research funding), the Doctoral Research Committee (Massey Doctoral Scholarship), and the Vic Davis Memorial Trust (Post-graduate Mental Health Practitioner Training Scholarship). Research and clinical colleagues have provided valuable suggestions, mentoring, and support over the last few years. Thank you particularly to John Fitzgerald, Simon Bennett, Don Baken, Ella Kahu, Jan Dickson, and Sarah Malthus. I would also like to thank Ben Ogles (developer of the Ohio Scales) for his generosity in allowing me to adapt his outcome measure for this research.

My appreciation for the unwavering encouragement, patience, and faith of my family and friends is huge. There are too many people to name specifically, but my special thanks go to Felicity, Josh, Mishka, Jake, Trudy, Sylvia, Sarah, Jo, Claire, Renee, Mel,

Di, John, Liesje, Laura, Rif, Melanie, Liz, Ana and Annette. I would like to especially acknowledge those who shared with me the internship, thesis retreat, and TAG meetings; you've been wonderful.

Finally, I am so grateful for the help of everyone around New Zealand who took part in this research. The High and Complex Needs Unit and the practitioners who participated in the survey or focus groups were so generous with their time. Thanks also to the school principals and teachers who kindly distributed the Ohio Scales. To the anonymous students and parents who completed those forms, thank you for your willingness to advance applied psychological research by sharing your personal information. Similarly, the carefully considered feedback on the Ohio Scales that was supplied by clinicians, parents, and young people was very much appreciated. I am so pleased and thankful that you helped fill the pages that follow with your contributions.

Thank you very much to everyone mentioned here, and to those who are not, but who were part of the process in small, but important ways.



AUTHOR'S NOTE

This thesis presents six studies on the topic of outcome measurement with young people in New Zealand. The first four studies were funded by a research contract with the High and Complex Needs Unit (HCN). *The HCN Outcome Measurement Project* was undertaken by a team of five clinical psychology researchers, of which I was a part. These studies were presented in a final report to HCN (Harvey, Barry, Fitzgerald, Evans, & Bennett, 2007), which is now publically available on their website. I would like to acknowledge the work of the team and make clear my contribution to the project and its connection to this thesis. It should be noted that from the outset of the project, it was agreed by the Principal Investigators (Prof. Evans and Dr Harvey) that the work I contributed would be considered suitable for submission as my doctoral thesis.

I was involved in all aspects of the project from writing the initial proposal to the oral presentation of the findings to HCN staff. While individual team members worked on specific studies, I was involved in all four, to varying degrees, as follows:

- *Practitioner Survey* – involved in the design of the survey and analysis of the data
- *HCN Advisor Focus Groups* – involved in the development of the questions and recruitment of participants, was one of three facilitators, conducted independent analysis of the transcripts
- *HCN Practitioner Focus Groups* – involved in the development of the questions and recruitment of participants, was one of two facilitators, conducted independent analysis of the transcripts
- *Measure Review* – (led this study) completed the search in its entirety, was one of two researchers screening the results, one of four reviewers of the final measures

In addition, the written work presented in the HCN report is not replicated exactly in this thesis. The chapters here were written by me using the data and analysis from the original project (in addition, the introductory chapters, the two additional Ohio studies, and the concluding chapter are my own work, independent of the HCN project/report). Parts of *Chapter 8* are drawn from a second contracted report completed for HCN (Barry, Baken, Fitzgerald, Evans, & Harvey, 2007), which is not publically available. Those parts were written by me originally, and small, relevant portions have been adapted for use in this thesis. The two reports referred to in this note are:

- Harvey, S., Barry, A., Fitzgerald, J., Evans, I., & Bennett, S. (2007). *Best practice outcome assessment for children and youth with high and complex needs*. Unpublished report for the High and Complex Needs Intersectoral Unit. Massey Psychology Clinic/The Psychology Centre. Palmerston North/Hamilton, NZ.
- Barry, A., Baken, D., Fitzgerald, J., Evans, I., & Harvey, S. (2007). *Preparing a formal outcome monitoring protocol for use with High and Complex Needs cases*. Unpublished report for the High and Complex Needs Intersectoral Unit. Massey University/The Psychology Centre. Palmerston North/Hamilton, NZ.

TABLE OF CONTENTS

ABSTRACT.....	I
ACKNOWLEDGMENTS	II
AUTHOR’S NOTE	IV
TABLE OF CONTENTS	VI
LIST OF TABLES	XIII
LIST OF FIGURES.....	XIV
LIST OF ABBREVIATIONS.....	XVI
CHAPTER 1 – OUTCOMES AND THEIR MEASUREMENT	1
<i>The rise of outcomes</i>	2
Early mental health services: Therapy is un-measurable.....	2
A shift: Symptoms and therapy are measurable.....	2
Economics and the provision of effective treatments	3
From effective services to effective people	4
<i>Outcome measurement</i>	5
Definitions	5

Outcome measures	6
The benefits of measuring outcomes	8
Research, policy, and funding.....	8
Organisation, agency, or service.....	10
Clinicians, young people, and their families	11
CHAPTER 2 – ROUTINE OUTCOME MEASUREMENT	15
<i>Routine outcome measurement</i>	<i>15</i>
<i>Requirements of a good outcome measure</i>	<i>16</i>
Psychometric profile	17
Usefulness.....	17
Feasibility.....	19
<i>Current practices</i>	<i>21</i>
International experience.....	21
United Kingdom	21
Australia.....	22
United States of America	25
New Zealand’s experience	27
MH-SMART and CAOS.....	27
Merry and colleagues’ report	29
Implementation of routine outcome measurement in New Zealand	31
<i>Introduction to the present research</i>	<i>32</i>
CHAPTER 3 – OUTCOME MEASUREMENT IN MULTIDISCIPLINARY SETTINGS	35
<i>Outcome measures for young people and multidisciplinary teams.....</i>	<i>35</i>
<i>The High and Complex Needs Strategy</i>	<i>36</i>
The High & Complex Needs Unit	37
HCN’s request for tenders	38
The HCN Outcome Measurement Project.....	40
<i>Introduction to the HCN studies</i>	<i>40</i>
CHAPTER 4 – STUDY ONE: CURRENT HCN MEASUREMENT PRACTICE.....	42
<i>Method.....</i>	<i>44</i>
Participants.....	44
Materials	45
Procedure.....	46
<i>Results</i>	<i>46</i>

Characteristics of participants and their clients.....	46
Participant characteristics	46
Client characteristics	47
Participants' current outcome measurement practice	49
Experience using outcome measures for HCN	49
Indicators to practitioners of client progress.....	51
Indicators to practitioners of client readiness to transition	52
Reasons for using outcome measures	53
Barriers to using outcome measures	54
Perceived usefulness of outcome measures	56
Ideal outcome measurement practice	57
Preparation required.....	57
Domains where measurement would be useful.....	59
Ideal approach	59
Recommended outcome measures.....	61
Additional issues.....	63
Cautions and reservations.....	63
Advice about measure selection	64
Advice about implementation.....	65
Summary	65
Characteristics of participants and their clients.....	65
Participants' current practice	66
Participants' ideal approach	68
Participants' recommendations for measures.....	68
Additional issues.....	69
Conclusions.....	70
CHAPTER 5 – STUDY TWO: CURRENT HCN MEASUREMENT PRACTICE	72
<i>Perspectives of HCN Advisors</i>	73
Method.....	73
Participants.....	73
Discussion questions	73
Procedure.....	74
Results	74
a. Current outcome measurement is problematic.....	75
b. Barriers to outcome measurement.....	76
c. Perceptions of the HCN teams' views	78
d. Ideal outcome measurement	78
Summary.....	80

<i>Perspectives of HCN practitioners</i>	82
Method.....	82
Participants.....	82
Discussion questions.....	83
Procedure.....	83
Results	83
a. Spotting change.....	84
b. Methods and tools.....	87
c. Accountability.....	89
d. Monitoring and management.....	90
e. Feedback from the client/family.....	91
Summary.....	91
<i>Conclusions</i>	92
CHAPTER 6 – STUDY THREE: THE SEARCH FOR SUITABLE MEASURES	95
<i>Parameters of the search</i>	95
The nature of the work of HCN	95
Domains of outcome.....	96
Criteria for the ideal measure.....	98
<i>The search for measures</i>	99
Identifying potential measures.....	99
Search of electronic sources.....	100
Search of catalogues	101
Search of published reviews.....	101
Screening the search results	102
The first screen	102
The second screen	103
The third screen.....	103
<i>Brief review of measures</i>	104
Measures of culture	105
1. Multigroup Ethnic Identity Measure.....	105
Measures of education.....	106
2. Canadian Achievement Tests – Third edition	106
Measures of family.....	107
3. Family Environment Scale.....	107
4. Index of Family Relations	108
5. McMaster Family Assessment Device – Version 3.....	109
6. North Carolina Family Assessment Scale – Version 2.0.....	110

7. Self-report Family Inventory	111
Measures of mental health	112
8. Hua Oranga	112
9. Pediatric Symptom Checklist.....	113
10. Strengths and Difficulties Questionnaire	114
Measures of physical health.....	115
11. Child Health and Illness Profile	115
Measures with several areas of focus.....	116
12. Behaviour Evaluation Scale – 3 rd edition	116
13. Child Health Questionnaire	117
14. Clinical Assessment Package for Risks and Strengths.....	118
15. Health of the Nation Outcome Scales for Children and Adolescents	119
16. Ohio Youth Problems, Functioning & Satisfaction Scales.....	120
17. Youth Outcome Questionnaire.....	121
Summary	122
CHAPTER 7 – STUDY FOUR: CONSULTATION AND SELECTION OF AN OUTCOME	
MEASURE.....	125
<i>Expert consultation</i>	<i>126</i>
Expert discussion of the short-listed measures.....	126
Expert recommendations	128
<i>Practitioner consultation</i>	<i>129</i>
Focus groups: Part II	129
Practitioner discussion of the short-listed measures	129
North Carolina Family Assessment Scale	130
The Ohio Youth Problems, Functioning, and Satisfaction Scales	131
Pediatric Symptom Checklist.....	133
Strengths and Difficulties Questionnaire.....	134
Youth Outcome Questionnaire	135
Practitioner recommendations	137
<i>Summary of expert & practitioner consultation</i>	<i>138</i>
<i>Selection of an appropriate outcome measure</i>	<i>139</i>
Issues to consider for implementing routine measurement	140
Final selection and recommendation	141
CHAPTER 8 – ADAPTING OVERSEAS MEASURES FOR LOCAL USE.....	142
<i>The Ohio Youth Problems, Functioning, & Satisfaction Scales</i>	<i>142</i>
Development and current use.....	142

Content and structure.....	143
Psychometric properties.....	145
<i>Preparation of measures for use in New Zealand.....</i>	146
Assessment of suitability.....	146
Preliminary local data.....	147
<i>Introduction to the Ohio Scales' studies.....</i>	147
CHAPTER 9 – STUDY FIVE: PRELIMINARY LOCAL DATA FOR THE OHIO SCALES.....	149
<i>Method.....</i>	150
Participants.....	150
Materials.....	151
Procedure.....	151
<i>Results.....</i>	152
Missing data.....	152
Scale performance.....	152
Reliability.....	152
Validity.....	153
Comparative analysis.....	153
Demographic information.....	153
Problem Severity and Functioning.....	154
<i>Summary.....</i>	157
CHAPTER 10 – STUDY SIX: INFORMANT FEEDBACK ON THE APPROPRIATENESS OF THE OHIO SCALES FOR NEW ZEALAND.....	160
<i>Method.....</i>	161
Participants.....	161
Workers.....	162
Parents/caregivers.....	162
Youth.....	162
Materials.....	162
Procedure.....	163
<i>Results.....</i>	163
Client details section.....	164
Problem Severity scale items.....	164
“ROLES” section.....	164
Hopefulness and Satisfaction scale items.....	165
Functioning scale items.....	165

Bicultural appropriateness	166
General feedback	166
<i>Summary</i>	167
CHAPTER 11 – SUMMARY AND CONCLUSIONS	170
<i>Summary of research findings</i>	170
Aim 1	170
Aim 2.....	172
Aim 3.....	173
<i>Discussion</i>	174
REFERENCES	180
APPENDIX A – STUDY ONE: HCN PRACTITIONER SURVEY	198
APPENDIX B – STUDY TWO: HCN FOCUS GROUPS	208
APPENDIX C – STUDY FOUR: EXPERT AND PRACTITIONER CONSULTATION	214
APPENDIX D – STUDY FIVE: COLLECTION OF NZ DATA FOR THE OHIO SCALES	238
APPENDIX E – STUDY SIX: INFORMANT FEEDBACK ON THE APPROPRIATENESS OF THE OHIO SCALES FOR NZ	246

LIST OF TABLES

Table 1	<i>Proportions of Participants* Stating Each Reason to Use/Not Use (a) Any of the Measures Identified, (b) HoNOS-CA, and (c) SDQ</i>	30
Table 2	<i>Professional Discipline of Survey Participants</i>	47
Table 3	<i>Characteristics of Participants' HCN Clients</i>	48
Table 4	<i>Proportion of Participants Indicating Outcome Measures Would be Useful in Each Domain of Functioning.....</i>	59
Table 5	<i>Participants' Comments about Outcome Measures They Recommended.</i>	61
Table 6	<i>HCN's Seven Domains of Well-being and Associated Key Areas of Outcome.....</i>	97
Table 7	<i>Keywords Used to Search Electronic Databases.....</i>	100
Table 8	<i>Search Results: Measures Remaining After the Screening Process</i>	104
Table 9	<i>Comparison of the 17 Reviewed Measures Against the Requirements of a "Good Routine Outcome Measure"</i>	124
Table 10	<i>Reliability Estimates Reported for the Ohio Scales' Parallel Informant Forms</i>	145
Table 11	<i>Internal Consistency Estimates for the Ohio Scales with a NZ Sample...</i>	153
Table 12	<i>Correlation Between Problem Severity and Functioning Scale Scores....</i>	153
Table 13	<i>Comparison of Gender and Age for the Ohio and NZ Samples.....</i>	154
Table 14	<i>Comparison of Mean Scale Scores for the Ohio and NZ Samples.....</i>	155

LIST OF FIGURES

- Figure 1.* Illustration of the increase in outcome literature by comparing the number of citations in a PsychInfo search, indexed on 1986..... 4
- Figure 2.* Proportion of participants rating the significance of each presenting issue for the HCN clients they have worked with. Rated on a scale from 1 (not significant at all) to 5 (very significant issue), and ranked top to bottom by mean rating..... 49
- Figure 3.* Proportion of participants rating the importance of each source of information for determining the progress of HCN clients they have worked with. Rated on a scale from 1 (not important at all) to 5 (very important indicator), and ranked top to bottom by mean rating. 51
- Figure 4.* Proportion of participants rating the importance of each source of information for determining whether HCN clients they have worked with are ready to transition back to mainstream services. Rated on a scale from 1 (not important at all) to 5 (very important indicator), and ranked top to bottom by mean rating. 53
- Figure 5.* Proportion of participants rating the importance of each reason why practitioners should use outcome measures. Rated on a scale from 1 (not important at all) to 5 (very important reason), and ranked top to bottom by mean rating..... 54
- Figure 6.* Proportion of participants rating the importance of each barrier to practitioners using outcome measures. Rated on a scale from 1 (not

important at all) to 5 (very important reason), and ranked top to bottom by mean rating..... 55

Figure 7. 95% confidence intervals for mean Problem Severity scale score for the three informant groups in the New Zealand and Ohio samples..... 156

Figure 8. 95% confidence intervals for mean Functioning scale score for the three informant groups in the New Zealand and Ohio samples..... 157

LIST OF ABBREVIATIONS

ABAS	Adaptive Behavior Assessment Scale
ACC	Accident & Compensation Corporation
BDI	Beck Depression Inventory
BES3	Behavior Evaluation Scale – 3 rd edition
BYIs	Beck Youth Inventories
CAMHS	Child & Adolescent Mental Health Services
CAOS	Classification & Outcomes Study
CASPARS	Clinical Assessment Package for Risks and Strengths
CAT3	Canadian Achievement Tests – 3 rd edition
CBCL	Child Behavior Checklist
CDI	Child Depression Inventory
CHIP	Child Health & Illness Profile
CHQ	Child Health Questionnaire
Conners'	Conners' Rating Scale-Revised
CORS	Child Outcome Rating Scale
CYF	Child, Youth & Family
DHB	District Health Board
DSM-IV	Diagnostic & Statistical Manual of Mental Disorders – 4 th edition
FACES	Family Adaptability & Cohesion Evaluation Scales
FAD	Family Assessment Device
FES	Family Environment Scale
GAF	Global Assessment of Functioning

HCN	High & Complex Needs Unit
HCN Strategy	Intersectoral Strategy for Children & Young People with High & Complex Needs
HoNOS	Health of the Nation Outcome Scales
HoNOS-CA	Health of the Nation Outcome Scales for Children & Adolescents
IFR	Index of Family Relations
IQ	Intelligence Quotient
KORS	Kaupapa Outcome Rating Scale
LSC	Local Service Coordinator
MEIM	Multigroup Ethnic Identity Measure
MH-SMART	Mental Health Standard Measures of Assessment & Recovery Initiative
MMYBO	Mental Measurements Yearbook Online
NCFAS	North Carolina Family Assessment Scale
NZ	New Zealand
Ohio Scales	The Ohio Youth Problems, Functioning, & Satisfaction Scales
ORS	Outcome Rating Scale
PSC	Pediatric Symptom Checklist
ROLES	Restrictiveness of Living Environments Scale
RTLB	Resource Teacher of Learning & Behaviour
SDQ	Strengths & Difficulties Questionnaire
SFI	Self-report Family Inventory
SRS	Session Rating Scale
SSA	Social Services Abstracts
TRF	Teacher Report Form
UK	United Kingdom
USA	United States of America
WISC	Wechsler Intelligence Scales for Children
YOQ	Youth Outcome Questionnaire
YSR	Youth Self Report

CHAPTER 1 – OUTCOMES AND THEIR MEASUREMENT

Clinicians working in the mental health field have experienced a powerful shift towards increased accountability in recent times. This shift has come from a variety of quarters, from international policy makers to researchers, clinicians, and clients (Slade, 2002). These multiple influences led to the evaluation of client outcome as one method of ensuring effective mental health services were being provided. There has been considerable debate about the value of evaluating outcomes and whether this should be a part of routine clinical practice (for example Blisker & Goldner, 2002; Callaly, Coombs, & Berk, 2003; Margison, et al., 2000; Trauer, 2003). However, many western countries have implemented mandatory procedures to evaluate outcome in mental health services, including New Zealand. As in other parts of the world, introducing these procedures has been challenging and met with some resistance from clinicians (Eggelston & Watkins, 2008), but offers a number of benefits that make it worthwhile.

The present research aimed to contribute a New Zealand perspective on outcome evaluation and to offer a potentially more acceptable measure to clinicians working with children, adolescents, and their families. This was sparked by the needs of a national multidisciplinary agency working with young people who have high and complex needs, who sought guidance on how best to evaluate outcome for their clients. To do this a series of studies were undertaken: practitioners were surveyed and invited to take part in focus groups, an extensive search of the international literature for potential

psychometric instruments was conducted, followed by expert review and practitioner feedback on the short-listed measures. Finally, the measure ultimately recommended was field-tested with a local sample, and feedback from informant groups was sought about the changes necessary for this overseas measure to be used in New Zealand. The research provides the first evidence-based recommendations for New Zealand on routine outcome measurement for use with young people in multidisciplinary settings, and builds on previous work showing the similarity of practitioner views on the topic here with those reported overseas. Before detailing each of these studies, an overview of the historical, international, and local context of “outcome” in mental health is presented by way of introduction.

The rise of outcomes

Early mental health services: Therapy is un-measurable

Large scale mental health services are a relatively recent phenomenon, generously funded by governments around the world since World War II (Speer, 1998). Speer commented that early on the “aura of mystique” (p. 2) that psychoanalysis gave these services provided protection from attempts to monitor symptoms or therapy because neither were conceptualised as measurable entities. In the 1950s, leading psychiatrists were invited to join expert researchers to prepare for the very first randomised controlled trial in psychiatry. As Tansella (2002) notes, “the psychiatrists immediately betrayed their scepticism by declaring the supremacy of knowledge acquired by doctors at the bedside and clinical intuition to be above any other methodology” (p. 4). This attitude may seem quaint to modern thinkers, but knowing this was in reference to a trial of psychiatric *medication* puts their notion of *psychotherapy* as un-measurable into a clearer context. In addition, the potent ingredient of therapy was seen as the relationship between analyst and patient, so introducing extra-treatment processes (such as outcome measurement) was seen to place the integrity of the treatment at risk (Lyons, Howard, O'Mahoney, & Lish, 1997). Ultimately, this meant that at that time measurement focused on *service* outcomes, such as the number of clients seen, length of treatment, number of sessions, and the cost involved (Ciarlo, 1982).

A shift: Symptoms and therapy are measurable

From the 1960-70s onward the combination of several factors led to a concerted push

for the measurement of *therapeutic* outcomes. First, debate raged within psychology following Hans Eysenck's (1952) conclusion that therapy was no more effective than the passage of time. This debate ignited a flurry of treatment-outcome research, which by 1970 showed psychological therapy was efficacious for treating mental illness (Lambert, Bergin, & Garfield, 2004). Second, there had been a concurrent increase in the popularity of behavioural and cognitive-behavioural therapies. These approaches challenged the prevailing psychoanalytic model by viewing therapy and symptoms as measurable phenomena (Speer, 1998). These two factors helped encourage the quantification of mental illness and therapy services, and this was new.

Economics and the provision of effective treatments

Further factors in the rise of outcome measurement were financial: government cost containment and the “managed care” movement. In the USA, restraining government spending was imperative to halt rapidly rising inflation. So in the 1980s, with health care costs at an all time high, mental health funding was in the spotlight (Eckert, 1994). Mental health services were no longer assumed to be helpful for clients; there was growing pressure for evidence they gave value for money. Added to this, and as part of the effort to contain costs, was the introduction of managed care organisations in the USA health system (Lyons, et al., 1997).

The managed care movement originally applied the model used for medical health to mental health, offering funding based on diagnosis. Because this resulted in inequitable resource allocation, this was later modified to emphasise outcomes instead, with the use of effective treatments being highlighted. This was consistent with the aim of managed care organisations to provide efficient and efficacious treatments (Lyons, et al., 1997); that is, the highest quality mental health care for the fewest resources. Similar developments were already underway in England and Australia (Slade, 2002). A substantially different view of service evaluation was evolving; captured succinctly by Ciarlo (1982), writing at the time,

...mental health program officials, sponsors, and other “stakeholders” are moving rapidly beyond simpler concerns about services delivered, target groups reached, and efficiency to the position that information on client outcome, or actual impact on clients treated, is essential to help justify why the programs are being mounted and sustained in the first place. (Emphasis in the original, p. 31)

From effective services to effective people

With funding depending increasingly on evidence of effective practices, research into how best to monitor client progress and measure therapy outcomes flourished. I conducted an update of Trauer's (1998) illustration of the expansion of the outcome literature. Replicating his methodology, two literature searches in the PsychInfo database were performed: one for the term *mental health* and one for the term *mental health and outcome*. The number of citations between 1986 and 2008 for each search term are shown in Figure 1 below, relative to their 1986 levels.

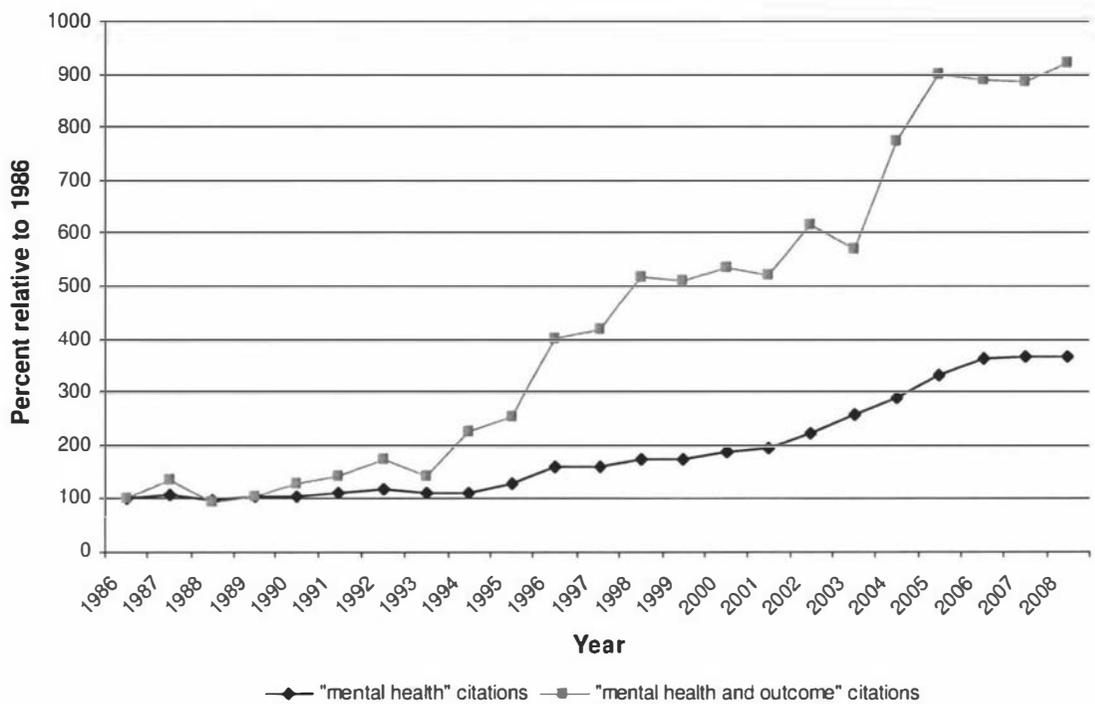


Figure 1. Illustration of the increase in outcome literature by comparing the number of citations in a PsychInfo search, indexed on 1986.

Although the citations for *mental health* have increased slowly since 1986, the citations for *mental health and outcome* increased at a significantly disproportionate rate, particularly during the 1990s. A partial correlation between year and *mental health and outcome* citations, controlling for *mental health* citations, was $r_s = .51$ ($df = 21$, $p = .003$).

This is not to suggest that it was only academics, researchers, and funders who were interested in the impact of services on client outcome. Friedman (1997) summarised the considerable change to the structure of child and family mental health services during this time and noted that these changes were led by clinicians. The innovative services that appeared then – where concepts such as, “systems of care”,

“wraparound”, “case management”, and “continuums of care” were conceived – reflected clinicians’ desire to provide effective services in a manner sensitive to the needs of individual clients (Lourie, 2003). Structuring services to address individual needs differed from the tradition of implementing a particular treatment model across a whole service, or targeting a particular diagnosis with a standard treatment, and it had consequences. One of these was the move from an exclusive focus on service outcomes to measuring individual therapy outcomes (Friedman, 1997). Lambert’s (2004) comment that “proof of effective treatment needs to be based on measurement of treatment response rather than provision of the ‘right’ treatment” (p. 9) sums up the thinking of these early innovative child and family practitioners. These new community-based and integrated systems of care led to renewed calls for evidence that they made efficient use of scarce mental health resources (Cross & Saxe, 1997; Wall, et al., 2005). They also sparked interest in the ability – or more accurately, the *lack* of ability – of services to provide good information to clinicians, children, and families about the impact of therapy (Huxley, 1994). Outcome measurement was, and is now, increasingly seen as a viable tool for addressing these concerns (Johnston & Gowers, 2005).

Outcome measurement

Definitions

Speer (1998) defines *outcomes* as “any characteristic of [clients’] behavior, condition, or adjustment that are significantly related to the reason for seeking mental health services” (p. 11). Trauer (1998) challenged the traditional notion of outcome from psychotherapy research, which was that client outcomes should be attributable to the intervention provided. Instead, he suggested a more pragmatic approach: considering outcome as the difference in the client’s status between two time points. This is more suitable for measuring change for clients in applied settings where identifying the active ingredients of an intervention is more difficult and less important than identifying improvements over time (Niumata-Faleafa & Lui, 2005). It is also important to note the difference intended in the present research between “goals” and “outcomes”. Goals are individualised predicted end-points a clinician hopes to achieve. “Outcomes” refers to what has actually been achieved, and may not represent an end-point, merely the status of the client at one time point relative to a previous point.

Measurement in therapeutic interventions is a method of summarising “what is comparable from one experience to the next” (Sonnanburg, 1996, p. 163). In addition, for the purposes of this research, measurement refers to “the use of a number to understand the frequency or intensity or duration” (Corcoran & Vandiver, 1996, p. 126) of those outcome characteristics. In this sense, *outcome measurement* is a method of summarising the change in a client’s status between two time points. It can also be used to aggregate the change of a number of clients to summarise outcomes for a clinician, for a service, or across demographic variables in a national system.

“Outcome measurement” in this research is used as a broad term to encompass the notions of both assessing outcomes and monitoring change over time. These are complementary processes, but have the potential to create difficulties in a practical sense (Fitzgerald, 2007). Assessing therapy outcomes can be done within a pre-post design (such as that used in treatment-outcome studies), and it is these data that are of interest to managers, funders, and policy makers, particularly when aggregated. Change during an intervention can also be monitored on an on-going basis and the data used to guide therapeutic efforts in real-time. Andrews and Page (2005) drew this distinction and highlighted the tension inherently created by the different purposes for which clinicians and management measure outcome. Although challenging, the studies in the present research sought to reduce this tension with a satisfactory compromise.

Outcome measures

As mentioned previously, the emergence of treatment-outcome research was one factor in the rise of outcome measurement. Researchers operationally define the variables they expect therapy to influence, and so develop methods of measuring change in those variables. These methods of measuring change in variables (or treatment outcomes) for research have offered a model and tools for clinicians to measure client treatment response in their own applied services (Corcoran & Vandiver, 1996).

Many approaches are available to evaluate whether any change has occurred during treatment, and the direction and extent of that change. Froyd, Lambert, and Froyd (1996) conducted a systematic, broad review of the outcome measurement strategies used in treatment-outcome research. They surveyed 21 journals over a five-year period and found 851 unique measurement strategies in over 300 studies. They

concluded that the most typical practice was to use a self-report pen-and-paper instrument for rating behaviour and symptoms. Psychometric instruments such as these have become almost synonymous with outcome measurement over recent decades and are the focus of the present research. For the purposes of this research, these instruments are referred to as *outcome measures* – provided of course they are designed with the intention to track client change and outcome (as opposed to being used for assessment or diagnostic purposes).

Lambert and colleagues (Lambert, Okiisji, Finch, & Johnson, 1998) predicted that by the year 2000 clinicians would be regularly using outcome measurement, which was said to add “a bit more science to their art” (p. 63). Unfortunately, this prediction was not quite borne out. In 2000, psychiatrists in the UK National Health Service reported low levels of routine measure¹ use for tracking change (measures for depression/anxiety used by 11.2% of psychiatrists; psychosis, 6.5%; cognitive impairment, 8.8%; substance use, 4.7%), and only 13.5% said their agency required routine use of standard measures (Gilbody, House, & Sheldon, 2002). A later survey of USA psychologists found only 37.1% used “some form” of outcome assessment (Hatfield & Ogles, 2004). Encouragingly, outcome measurement was noted to be more prevalent for those working with children and families; over half reported using outcome measures compared to barely a third of those working with adults.

However, it is important to note that many measures have often not been developed with applied settings in mind, although many lend themselves to this purpose. The almost overwhelming number of measures developed for research purposes address a multitude of outcome domains (Hill & Lambert, 2004). With such a plethora of measures available, clinicians face a daunting task when selecting a measure for their applied practice. Another finding of Hatfield and Ogles' (2004) survey, was that over 150 different measures were mentioned by participants, with 125 of those named by only one participant each. Naturally, they concluded such “measurement chaos” (p. 489) presented a challenge for those wishing to find a suitable measure for their practice. In recent years there have been measures purposefully developed for use in “real world” services, rather than formal research settings (for example, the “Ohio Scales”, Strengths and Difficulties Questionnaire, or Youth Outcome Questionnaire; Burlingame, Wells, & Lambert, 1996; Gowers, Harrington, Whitton, Lelliott, et al., 1999;

¹ For the sake of brevity, the term “measure(s)” used throughout can be taken to mean “outcome measure(s)”, unless otherwise indicated.

Ogles, Melendez, Davis, & Lunnen, 1999). This move reflects a growing appreciation of the benefits of using structured approaches to monitoring outcome in applied clinical settings.

The benefits of measuring outcomes

The literature supporting outcome measurement is founded on the assumption that it is a useful practice. Certainly, in fields such as business, education, and medicine, utilising outcome information about customers, students, and patients is well established. In mental health, one difficulty is that outcome information is not always as obvious, concrete, or accessible. For example, blood sugar tests are an obvious indicator of how well diabetes is being managed, NCEA² results are a concrete method of determining the extent of high school students' knowledge, and sales figures to gauge the success of a new product launch are readily available. Fewer of these observable and easily available indicators arise as part of "business as usual" in mental health, aside from service outcomes, such as length of treatment, which are of dubious value as predictors of improved mental health status (Goldney, Fisher, & Walmsley, 1998; Noser & Bickman, 2000).

In order to gain from outcome feedback in the way other fields do, mental health services must actively collect and use outcome information. This is likely to increasingly occur as the value of measuring outcomes becomes more widely appreciated (Hatfield & Ogles, 2004; Lambert, Okiisji, et al., 1998). The numerous benefits to investing time and effort to measure progress and outcome are grouped here as being evident at three levels: (1) research, policy, and funding, (2) organisation, agency, or service, and perhaps most importantly, (3) clinicians, young people, and their families.

Research, policy, and funding

In the past, clients were expected to utilise (and governments expected to fund) mental health services because they were seen as intrinsically valuable, not because there was empirical evidence of their quality (Noser & Bickman, 2000). However, this assumption has been especially challenged in the last two decades (Lambert, Salzer, & Bickman, 1998), and the public and funders are seeking to make more informed decisions about mental health treatment. Bickman (2008) captures the situation well,

² The National Certificate in Educational Achievement (NCEA) is New Zealand's national high school qualification offered at three levels for Year 11, 12, and 13 students.

If you were in the market for a car, would you buy it based on the following information: It is located on 5th and Main, costs \$20,000, is 2 years old, is a Ford, and seats six? Whenever I give that example in a talk I don't get a single potential buyer...Now let's look at what we know about mental health services. We know the location, the charges, how long the organisation has been in existence, whether it is inpatient, day treatment or outpatient, and just maybe how many clients it serves. Like the car, we know nothing about its quality, performance, or durability. Yet this is how we buy and sell mental health services. (p. 437)

Parents are able to make an informed choice about the school their child attends because information about quality, curriculum, and outcomes are available. For young people and their families who seek assistance from mental health services, often in highly distressing and pressured times, an equally informed choice is not possible. Being able to openly talk with clients about their choices and expectations of treatment is an ethical imperative (Code of Ethics Review Group, 2002) so as to avoid inadvertent exploitation of those who are desperate for help. To enable this type of openness and informed decision-making, information about the effectiveness of real world services for children and families needs to be readily available (Brann & Coleman, 2010; Te Pou, 2009).

Researchers and academics have called for more data to be collected about the effect of treatment in applied settings, due to the difficulty replicating promising treatment efficacy results from clinical trials (Andrews, 1999). This would allow better understanding of what works outside these carefully controlled studies and in turn would advance the evidence-based practice literature that clinicians are encouraged to depend upon for treatment planning. Applied outcome data would also supply the necessary detail for further investigation of apparently ineffective services using alternative designs such as the feedback research approach (see Rog & Bickman, 1984) or the learning organisation model (for example Birlleson, 1998). Analysis in this way would give researchers and services suggestions about possible future directions for improvement at a local level.

Given the increasing emphasis on obtaining value for money, tax payers, government, and other funders are also interested in client outcome (Johnston & Gowers, 2005). There are frequent calls from clinicians and the public for more resources to be directed to mental health and these are likely to be met with resistance unless evidence

shows this is a good investment (Andrews, 1999). Aggregated outcome data could be one source of evidence to show change for clients commensurate with resource investment. Combined with research showing that real world services are effective, public perception of the value of psychological therapy could also be improved. On the other hand, perhaps aggregated outcome data might show minimal positive effects for clients. While unpalatable to most stakeholders, this is also valuable information if the aim is to provide useful services. Outcome measurement should not be avoided purely to provide protection from the truth that applied services are ineffective, if that really is the case (Bickman, 2008).

On a more positive note, there is evidence that the use of outcome measures leads to more efficient use of therapist and client time (Lambert, et al., 2003; Lambert, et al., 2001). When therapists received feedback (from outcome measures) for clients who were “not on track”, they were seen for almost twice as long as when therapists did not receive feedback, and had better outcomes. Conversely, when therapists received feedback for clients who were “on track”, fewer sessions were used than when therapists did not have feedback. The interesting proposition for service providers and funders is that there was an *overall reduction* in mean treatment duration when therapists received feedback, which means lower costs for improved client outcomes. With research findings such as this, it is easy to see why those at the policy and funding level in mental health would encourage outcome measurement.

Organisation, agency, or service

In a summary of the development of child and adolescent mental health services (CAMHS) in the UK, Cottrell and Kraam (2005) predicted increasing use of evidence-based treatment approaches in the future. However, they noted the absence of evidence for the effectiveness of treatments for the complexities of real world clients (see also, The Werry Centre, 2010) or the new manner in which CAMHS are increasingly being structured (e.g. multidisciplinary teams, increased community liaison, more inter-agency collaboration). They also mentioned the lack of evidence about effectiveness for each profession (i.e. psychologists, social workers, psychiatrists, nurses etc.). Outcome data would enable services to begin to collate evidence regarding how treatments, structures, and professions are working *in their own organisation*. This would assist managers at a local level to make informed decisions about changes to processes and structures to improve the quality of care they provide.

In addition, outcome measurement could allow District Health Boards (DHBs) to compare their outcomes against national benchmarks and with each other (Kriebler, 2003). Hodges (2004) reported the experience in Michigan where this happens between services throughout the State. She noted that clinicians were energised by having this comparative feedback, which was a motivator to improve services for the client groups for whom their outcomes were poorest.

Clinicians, young people, and their families

From the clinician's perspective, the main benefits of outcome measurement are treatment and client-related, rather than those associated with third parties (Hatfield & Ogles, 2007). The three most important reasons psychologists use outcome measures are, (1) to track client progress, (2) to determine the need to alter treatment, and (3) because it is part of ethical practice (Hatfield & Ogles, 2004). Lambert and Hawkins (2004) also felt monitoring client progress was part of good clinical and ethical practice. Outcome measurement is one method for doing this monitoring, but also provides an opportunity for clinicians to obtain external information about their effectiveness, rather than relying upon their own, perhaps biased, appraisal (Hazelton & Farrell, 1998). This feedback on their practice is information valued by clinicians (Bickman, et al., 2000). Indeed, Miller, Duncan, Brown, Sorrell, and Chalk (2006) have gone so far as to call for the use of outcome measurement as a means of establishing the effectiveness of clinicians, rather than the effectiveness of services or treatment models:

...rather than evidence-based practice, therapists tailor services to the individual client via practice-based evidence; instead of empirically supported therapies, consumers would have access to empirically validated therapists. (Emphasis added, p. 17)

A number of studies have looked at providing outcome measure feedback to clinicians about clients' progress in therapy, with the consensus that this has positive effects. Based on their series of well-designed studies, Lambert, Harmon, Slade, Whipple, and Hawkins (2005) reported clients whose therapist had feedback about their progress achieved the same significant level of improvement sooner than clients whose therapists did not have such information. Also, supplying therapists with feedback and a decision-making tool (essentially a flow chart of options) led to better client outcomes than supplying feedback alone (Whipple, et al., 2003). Miller, Duncan, Brown, Sorrell, and Chalk (2006) reported the striking finding from a two-year study with well over 6000 clients, that therapists receiving information from outcome measures "effectively doubled the effect size of services" (p. 14) from 0.37 at baseline (measures completed,

but results not provided to therapists) to 0.79 at the end of the study. They then removed the data of clients whose intake scores were not in the clinical range, and the analysis showed the final effect size increased to 1.06. In addition, using a measure to monitor the therapeutic relationship was also found to lead to better client outcome. These studies suggest that in order to obtain the most benefit from outcome measures, clinicians need to *use* the information rather than simply collect it.

There are a number of ways this outcome information could be used by clinicians. One might be to flag cases at risk of a poor treatment outcome. Sapyta, Riemer, and Bickman (2005) conducted a meta-analysis of randomised controlled trials of feedback to clinicians about client health status during treatment. They found clients in feedback conditions had significantly improved outcomes compared to controls (mean effect size, 0.21), with more benefit for clients who were not progressing well (mean effect size, 0.31). Given that therapists' ability to identify clients at risk of negative treatment outcome is generally poor (Hannan, et al., 2005; Lambert & Ogles, 2004), this important finding suggests measures could have a role in flagging these cases so clinicians can re-evaluate the treatment plan much earlier on. In fact, studies have shown that feedback from measures about cases predicted to fail not only enhanced outcome compared to controls, but also reduced drop-out and deterioration rates (Lambert, et al., 2001; Lambert, et al., 2002). These studies have been conducted for outcomes feedback to therapists working with adult clients. An early warning system based on the work of Lambert and colleagues has been tested for use with the "Ohio Scales", a measure used by therapists for children and adolescents. Using outcome data for over 20 000 CAMHS clients, the "Problem Severity Scale" of the measure identified cases at risk of treatment failure with an overall hit rate of 95% (Healy & Tam, 2009). Given that current estimates are that 15-25% of young people (compared with only 5-10% of adults) deteriorate in treatment (Lambert, 2010), the development of warning systems using outcome measures is particularly pertinent for child therapists.

Clinical decision-making can also be enhanced by the use of outcome information. Hatfield and Ogles' (2006) empirical investigation showed that despite clinicians' assertion that client verbal report was more influential for their decision-making than information from measures, they actually used both sources equally when judging client progress. However, they more often chose to alter their treatment plan after measures revealed deterioration than when deterioration was verbally reported by the client. Taken together, these findings indicated that using measures could lead to more timely treatment modifications for clients who are not improving as expected.

Clinicians receiving feedback do not appear, over time, to get better at identifying clients at risk of negative outcome (Lambert, 2003, cited in Miller, Duncan, Sorrell, & Brown, 2005), therefore measures have an important role to play in facilitating clinical decision-making.

In practical terms, using measures is likely to bring to both the client and clinician a focus on goals and re-evaluation of the treatment plan (Lambert, Okiisji, et al., 1998; Slonim-Nevo & Anson, 1998). A recent study supported this, proposing that the benefits of using *DIALOG* (a regular, structured measure of the client's perspective in eight life domains and three treatment domains) were the result of increased clinician-client communication and therefore better focussing of sessions (Bylund, 2008). A randomised controlled trial design was used to compare usual care alone, with usual care plus bi-monthly *DIALOG* (Priebe, et al., 2007). After 12 months, compared to those in the usual care group, clients in the *DIALOG* group had significantly improved quality of life and satisfaction with treatment, and less unmet need. The increased communication and session focus that measures afford is valuable, as client-clinician agreement on goals is consistently found to be positively associated with outcome (Orlinsky, Ronnestad, & Willutzki, 2004) as shown in a recent meta-analysis that reported an effect size of .34 for goal consensus and outcome (Tryon & Winograd, 2011).

Almost all research into the effect of using outcome measure information has been designed with feedback to the therapist in mind. However, 98% of clients would like to access this outcome information (Harmon, Hawkins, Lambert, Slade, & Whipple, 2005). There is little published on the effect of supplying clients with feedback on their progress. Lambert and colleagues' (2005) study comparing multiple therapist feedback conditions and treatment-as-usual reported the highest rate of improvement and lowest rate of deterioration were when feedback was also given to the client (more than double that of treatment-as-usual). A review and meta-analysis of outcome feedback suggested that due to the small number of studies on feedback to clients, although results are promising, further work is required (Lambert & Shimokawa, 2011).

Following their series of studies (including Harmon, et al., 2005; Lambert, et al., 2003; Lambert, et al., 2001; Lambert, et al., 2002; Whipple, et al., 2003), these researchers recommended the widespread introduction of feedback on client outcome in mental health services. They specifically drew attention to how effectively clinicians have been encouraged to use empirically-supported treatments based on small effect sizes (0-

0.20), and how this compares poorly to using outcomes feedback where effects are much larger, from 0.34-0.92 (Lambert, et al., 2005). The evidence is building to support these recommendations, because by using feedback from outcome measures, clinicians can enhance the effectiveness of their interventions and improve outcomes for clients.

Outcome measurement started out as a method for governments and funders to increase accountability and encourage better value for money, but is now recommended as good clinical practice in itself (e.g. Hazelton & Farrell, 1998; Johnston & Gowers, 2005; Lambert & Hawkins, 2004; Margison, et al., 2000). There are benefits for researchers, policy makers and funders, services and their management, clinicians and their clients from investing in collecting and using the information from outcome measures (Trauer, 2010). Many of these benefits are enhanced when measures are used by clinicians across a service on a routine basis. This is the topic of the next chapter.

CHAPTER 2 – ROUTINE OUTCOME MEASUREMENT

As Newnham and Page (2007) remark, monitoring client progress is “a win-win proposition for clinicians, health-care organisations and clients who all share the same goal” (p.4). Recognising the multiple benefits, many public mental health services around the world have introduced routine outcome measurement. These efforts, including the New Zealand experience, are described shortly, along with an outline of the potential contribution of the present research to the current knowledge base. First though, the notion of routine outcome measurement and the requirements of a good routine outcome measure are explained.

Routine outcome measurement

There is a difference between outcome measurement in research to generate new knowledge and outcome measurement in applied settings to monitor client progress on a routine basis (Spence, Donald, Dower, Woodward, & Lacherez, 2002). However, applied clinicians look to research to guide their selection of a measure to use with clients. Due to the chaotic array of measures used in treatment-outcome research, Froyd, Lambert, and Froyd (1996) suggested an urgent coordinated effort by researchers to agree upon the “best” measures. This, they said, would reduce the likelihood of individual clinicians and applied services using a similarly diverse range of measurement strategies. Yet, the attempts to identify the best measures have really

been driven more by mental health services than by researchers (for example, Bickman, et al., 1998; Merry, et al., 2004).

Mental health services have directed their efforts at identifying a measure that can be used in a standard manner by all clinicians, with all clients, across time, on a regular basis. This idea of *routine outcome measurement* locates the evaluation of progress and outcome as part of usual practice in the service; it becomes a routine activity. The emphasis is on the use of a standard measure, or set of measures, that are used regularly during each client's time with the service. Not surprisingly, there have been critics of mandatory routine outcome measurement (such as, Blisker & Goldner, 2002; Lakeman, 2004) and these criticisms have been crucial for guiding the discussion of the requirements of measures for routine use in a variety of international services.

Requirements of a good outcome measure

The field of outcome measurement has become more applied as it has matured, but not necessarily more coherent (Hill & Lambert, 2004). The treatment-outcome literature that was useful as a source of new instruments, has not been as helpful in providing guidance for services trying to choose among them. In light of this lack of direction from research about what to use in practice, there have been a multitude of frameworks, guidelines, best practice models, and selection criteria produced (for example, Friesen & Winters, 2003; Hill & Lambert, 2004; Kazdin, 2004; Speer, 1998; Spence, et al., 2002). This has been helpful for refining the field of outcome measurement and provides a good base for projects such as this, which involve a search for measures.

The literature has turned to address the refinement of routine outcome measurement practices. Two major areas of concern have been about what content measures ought to address and what criteria ought to guide the selection of a routine measure (Rock, Combrinck, & Groves, 2001). Although there has been progress, these matters continue to be debated (for example, at the inaugural *He Kakano: Australasian Mental Health Outcomes Conference* in 2007). There are consistencies however, and the requirements for a good routine outcome measure are loosely grouped here as concerning (1) the psychometric profile, (2) its usefulness to those interested in the results, and (3) its feasibility in applied practice.

Psychometric profile

Clearly, any routine outcome measure must measure what it claims to and do so in a manner that produces reliable results that assist clinical decision-making. Cicchetti (1994) provided guidelines for assessing the reliability coefficients of psychometric measures in psychology as follows:

- Poor Below .40
- Fair .40 – .59
- Good .60 – .74
- Excellent .75 – 1.00

A measure needs to have adequate internal consistency and inter-rater reliability and be sensitive to change if it is to have utility as a monitoring tool. Very high test-retest reliability is usually sought when the constructs being measured are understood as being static or unchanging over time (IQ or personality traits, for example), but may not be desirable in an outcome measure (Nixon, 1997). Test-retest reliability is expected to be slightly lower for outcome measures as they should tap changeable domains and detect shifts in these over short intervals. Internal consistency estimates are therefore considered more important than test-retest reliability when selecting an outcome measure (Groth-Marnat, 2003). The structure of the measure can also impact on how sensitive to change it is (Coolican, 1999). For example, items rated as “true/false” are not likely to detect change as readily as items rated on a multi-point continuum or Likert-type or interval scales.

The measure should have adequate face validity in order to engage completers and instil confidence in its ability to represent the client’s current situation sufficiently (Spence, et al., 2002). There should be evidence of formal validity checks, such as comparison with scores on other concurrent measures (divergent and convergent validity) and with objective indicators of progress (criterion validity). This assures the user that its data represent some meaningful indication of outcome. Finally, the measure needs to be valid for the population it is intended to be used with (Geisinger, 1994).

Usefulness

A good routine outcome measure must be useful to those who are interested in the information it provides. Therefore, the content of the measure and individual items should be applicable to the purposes and needs of the service or clinician using it.

Clinicians view improved functioning and symptom reduction as the two most important areas of outcome to be assessed (Bickman, et al., 2000; Garland, Kruse, & Aarons, 2003). Speer (1998) summarised the essential content areas that had consensus literature support as being distress, symptoms, and functioning, but conceded these often overlap. Although considering it important, he did not view client satisfaction as being an indicator of outcome. In contrast, later authors regard satisfaction as a crucial area to be measured; one which is of high importance to young people and their families (Friesen & Winters, 2003; Ruggeri, 2001). Plante, Couchman and Hoffman (1998) noted that even when measures show no significant improvement in symptoms, having data about satisfaction is vital for providing a complete picture. They suggest that high client satisfaction with no concomitant change in symptoms could indicate that “important ‘care’ (as compared with ‘cure’) is occurring” (p. 54). New Zealand mental health policy has encouraged inclusion of the client’s perspective (Kriebler, 2003), so measuring satisfaction would be important. Much of the literature advocates measuring a much broader range of content than that proposed by Speer. Specifically, a measure should ideally cover symptoms, functioning, satisfaction, and quality of life (Koch, Lewis, & McCall, 1998).

Another content issue regards whose perspective on progress and outcome is obtained. Multiple systemic factors impact on the well-being of young people and a good outcome measure must take into account the family, developmental, educational and social features of their presentation (Hunter, Higginson, & Garralda, 1996). More so than in adult services, collateral information from parents or others is a vital addition to the clinician’s perspective. Consequently, measures should allow for multiple perspectives on the case and using parallel informant forms is one way to do this. Melendez (2003) investigated the differential effect of outcome measure feedback to the clinician for young people when a single clinician-rated measure was used compared to when parallel measures were used. Although positive outcomes were noted for both groups, better outcomes were observed in the multiple informant condition. Hunter and colleagues (1996) suggested the view of the young person, parent(s), and teacher are needed, and on parallel forms this information can be triangulated for validity or therapeutic purposes (De Los Reyes & Kazdin, 2006; Trauer, 2003).

Although being able to compute an improvement in raw scores on a measure over time could be useful, there are other methods of making use of the scores. First, with all things equal, a specific outcome measure (i.e., with a range of different content areas)

is preferable to a global measure (Carr, 2006; Phelan, Wykes, & Goldman, 2001). For example, it is more useful to see no change on a symptom subscale and improvement on a functioning or quality of life subscale than to see any of these results in isolation or “lost” in a combined global score. Second, it is useful for measures to include clinical cut-off scores or descriptive ranges (e.g., mild, moderate, severe). These allow the clinician to interpret the client’s score in a meaningful way and track movement toward the non-clinical range, which could inform decisions about ending treatment. Finally, a means to standardise the raw score using a normative comparison is ideal (Kendall & Sheldrick, 2000). Normative data may be from a clinical or community population, or both, and include sub-groups such as gender or education level. With children and adolescents it is particularly important that normative data reflects age/developmental differences (Spence, et al., 2002) and this could be done with age-adjusted scores or norms grouped for specific age ranges. Again, normative comparisons allow meaningful interpretation of a raw score and therefore enhance the utility of the measure (Eatwell & Wilson, 2007).

Ideally, a routine outcome measure would be useful at multiple levels (Te Pou, 2009). This could mean provision of data for epidemiological or population level purposes to be used for research, benchmarking, or policy development. Data at a local organisation or service level would enable performance and effectiveness audits, inform management decision-making, and facilitate quality improvement. Most importantly, the measure must be useful at the clinical level. The individual clinician and client must find that the completion and interpretation of the measure adds to the therapeutic work they do together. If this is not the case, the measure will be in danger of falling into disuse, which deprives all other levels of the potential benefits routine outcome measurement offers (Te Pou, 2009).

Feasibility

Brief administration time and ease of use are frequently cited as critical in assessing the feasibility of a measure (Hunter, et al., 1996). Perhaps even more so for clinicians working with young people when multiple parallel forms are administered, scored and interpreted. For busy clinicians with large caseloads, administering measures on a regular basis should take 15 minutes or less (Groth-Marnat, 2003).

Routine measurement in settings where a number of disciplines work with young people requires a measure that can be used by all clinicians. Many psychometric

instruments can only be purchased and used in New Zealand by those with advanced training in psychometric assessment (i.e., Level C Test User Registration with the New Zealand Council for Educational Research). Obviously, this is a barrier to routine outcome measurement as many multidisciplinary practitioners are not eligible for registration to use measures at this level. Therefore, a good routine outcome measure should be able to be administered with little training, and be interpreted by clinicians with a range of professional experience.

Another practical consideration is that measures be financially viable. This applies to both the cost of the measure and the resources for implementation and ongoing use. This might include training clinicians and obtaining the technical capacity required (particularly for computer-based or other technology-supported measures). With a shortage of mental health funding, measures which are in the public domain or available at minimal expense are preferable.

Feasibility is a vital consideration when selecting a measure because user-friendliness is likely to influence whether it is acceptable to clinicians and clients. As obtaining the “buy-in” of these groups is essential to make routine measurement work (Bickman, et al., 1998), feasibility may outweigh other features of an otherwise suitable measure.

To summarise, a good routine outcome measure for use with young people, ideally, is required to,

- Be reliable and valid; psychometrically robust.
- Be sensitive to change.
- Have items that are applicable to the needs of the client group.
- Cover symptoms, functioning, satisfaction, and quality of life.
- Consider family, developmental, educational and social aspects of well-being.
- Include the perspectives of the clinician, parent, young person, teacher or other important adults.
- Have cut-off scores/descriptive ranges and normative data.
- Produce results that are useful at a clinical level primarily, but at management and research/policy levels too.
- Take no more than 15 minutes to complete.
- Need little training to administer and interpret.
- Be available for minimal cost.
- Be user-friendly.

Current practices

International experience

Services for children and families around the world have commissioned research to inform the implementation of routine outcome measurement. The experiences of services from the UK, USA, and Australia, who have all introduced routine protocols in varying forms, highlight consistent issues in two main areas: choosing a measure that meets the requirements outlined above and overcoming resistance from clinicians to the notion of routine monitoring.

United Kingdom

In the 1990s the UK Department of Health proposed to quantify and measure the outcomes of mental health clients in the public system and after finding nothing suitable already existing, a new measure was developed (Wing, et al., 1998). The clinician-rated *Health of the Nation Outcome Scale* (HoNOS; Wing, et al., 1998) was designed to address gaps left by contemporary measures in adult mental health; that is, to be “brief enough for routine use across the country while also covering the clinical and social range required” (p. 11). Wing et al. concluded from their final field-test that HoNOS was simple, “generally clinically acceptable” (p. 16), had broad coverage, and adequate psychometrics. They advocated its widespread routine use for tracking individual client progress and for national administrative and epidemiological uses.

Later research was not quite so optimistic, with concerns about the reliability and specificity of items, clinicians’ ability to understand them and make ratings, and the exclusion of the client’s perspective (Orrell, Yard, Handysides, & Schapira, 1999). Others concluded that although HoNOS might be useful for tracking changes in social functioning (Bebbington, Brugha, Hill, Marsden, & Window, 1999; Slade, Beck, Bindman, Thornicroft, & Wright, 1999), or affective and psychotic disorders (Amin, et al., 1999), and was practically feasible (Sharma, Wilkinson, & Fear, 1999), it was unsuitable for treatment planning (Sharma, et al., 1999; Slade, et al., 1999), and was “inadequate as a measure of symptoms” (Bebbington, et al., 1999, p. 391). Therefore, it was thought there were “serious problems” (Bebbington, et al., 1999, p. 394) with using HoNOS as a routine measure of national mental health outcomes and its “widespread adoption...for use in routine clinical practice would be premature” (Sharma, et al., 1999, p. 395). In summarising these findings at the time, Stein (1999) predicted HoNOS would not become part of routine clinical practice and questioned

whether it could improve the mental health of clients in the UK or any other country. Given their similar format, many of these criticisms are also applicable to its companion measure for children and adolescents, which clinicians were cautious about embracing (Jaffa, 2000) given its limitations (Garraalda & Yates, 2000).

The *Health of the Nation Scales for Children and Adolescents* (HoNOS-CA; Gowers, Harrington, Whitton, Lelliott, et al., 1999) was developed as part of the HoNOS suite of lifespan outcome measures (which includes measures for adults, older adults, and other specific groups). As for adult mental health, a literature review revealed no existing suitable measures for routine use in the UK (Hunter, et al., 1996) and HoNOS-CA was seen as having the potential to address the general absence of purpose-designed outcome measures in the child and adolescent field. Gowers and colleagues (1999) reported adequate reliability, validity, and sensitivity to change and suggested it would be a useful tool for UK CAMHS to use on a routine basis. They also reported that most CAMHS staff (87%) involved in their research thought the measure was easy to use. However, fewer staff felt HoNOS-CA was *useful* for routine practice; the authors stated vaguely “more than half the respondents” (p. 414).

Although not mandated, a selection of measures (including HoNOS-CA) was suggested for routine use by UK CAMHS, but Johnston and Gowers' survey (2005) showed less than 39% of CAMHS used HoNOS-CA and around 56% used the *Strengths and Difficulties Questionnaire*. The SDQ (Goodman, 2001) was most commonly used, followed closely by the *Conners' Rating Scales-Revised* (Conners, 1997) and the survey also found that services certainly were interested in collecting outcome data. The confusing state of the child and adolescent outcome measurement literature may explain the finding that 46% of those collecting data used measures no other CAMHS was using, or had developed their own. More encouraging was that three quarters of services measuring outcome on a routine basis collected baseline, during treatment, and discharge data, which managers said was for client monitoring purposes. However, fewer services used the data for feedback to clients (47%) or clinicians (38%), which is disappointing given the potential benefits this could have.

Australia

Australia has had a mounting focus on quality of client care and outcome since the early 1990s (Slade, 2002). They were the first to introduce a truly national routine outcome measurement protocol in public mental health systems, and to date, New Zealand is the only other country to have done this (Trauer, 2010). Like the UK, after

reviewing the literature and field-testing outcome measures for adults (Andrews, Peters, & Teesson, 1994), they turned to the child and adolescent field. Bickman and colleagues (1998) were commissioned to research and recommend a way forward for public CAMHS nationwide. Their comprehensive review of over 180 measures found none was ideal for immediate implementation; however, they outlined a plan to move services toward routine outcome data collection. This plan included the introduction of three “modules” designed to collect comprehensive information about the progress of a case through the CAMHS system. One module required new measures be developed, but for the other two modules the *Ohio Youth Problems, Functioning and Satisfaction Scales* (the “Ohio Scales”; Ogles, et al., 1999) was recommended as the core outcome measure. At that time, the Ohio Scales had only recently been developed and was still being refined. Despite this, Bickman et al. felt the measure had sufficient potential to recommend investment in its development for national use. This recommendation was not taken up and Australia mandated the use of two measures: the HoNOS-CA and SDQ; rated by the clinician and parents/young person, respectively. Both measures had been considered by Bickman et al., but because they were still being validated did not meet the selection criteria for their critical review.

Brann, Coleman, and Luk (2001) later evaluated the performance of HoNOS-CA in multidisciplinary CAMHS services in Melbourne. A strength of this work was its real world clinical setting, including brief clinician training, which is typical for applied practice. The HoNOS-CA completion rate was 87% and although they found it to be used reliably by each rater, there was only moderate reliability when the raw scores were compared across raters. This is important because of the implications for the utility of the measure. It suggests a clinician using HoNOS-CA to monitor his or her own client’s progress over time could do so with a degree of confidence. However, if the same client is being rated by different clinicians at different times, if the service examines aggregated HoNOS-CA data, or cut-off scores are being determined, perhaps a more cautious interpretation of the data would be warranted. Brann et al. (2001) also found the improvement in scores across a three month period was significantly correlated with the clinician’s perception of client progress. This is not surprising since the HoNOS-CA score is essentially a quantification of the clinician’s impression of the client’s problems. Although they did not compare to other indicators of improvement, there was no significant difference between the scores of those who had been discharged and those still being treated at the three-month point in the study, which possibly indicated a lack of sensitivity to change.

A similar project was completed with adult mental health services across four states using the mandated HoNOS (Goldney, et al., 1998). The completion rates varied from 41-82% across hospitals. The authors noted higher rates were for hospitals who regarded HoNOS as part of good practice rather than simply a research exercise, and the increased paperwork was thought to be another factor in the lower rates. There was no significant relationship between total score and length of stay, overall or for specific diagnostic groups. This prompted the authors to recommend caution when considering the HoNOS for resource allocation purposes. The strongest relationship to length of stay was the “Problems of Daily Living” scale, which lends support to the notion of tracking functioning in addition to symptoms.

There have been a few articles addressing routine outcome measurement experiences in Australia, some sparking considerable debate. For example, Lakeman (2004) responded to an editorial (Coombs & Meehan, 2003) which enthusiastically explained the introduction of routine measurement in Australia. Lakeman urged mental health nurses to be wary about the introduction of the mandated measures. His critique succinctly captured the concerns many have raised elsewhere (for example, Blisker & Goldner, 2002; Happell, 2008), and this reinforces the need to ensure clinician buy-in in order for routine measurement to be implemented successfully (Rock, et al., 2001). To summarise, Lakeman’s concerns were:

- Measures cannot capture the important individual differences of a client as *outcome* means different things to different stakeholders.
- Ratings are influenced by multiple factors and the validity of the mandated measures was questioned.
- Measures do not monitor the therapeutic relationship, which is central to nurses’ work.
- The simplistic application of routine outcome measurement is expensive and has no benefits for the psychosocial outcome of clients.
- Measures are reductionist and strip the client’s experience of all meaning.
- Routine outcome measurement “draws nurses into a potential web of collusion with psychiatric oppression” (p. 213), which is not ethical or consistent with a recovery orientation.
- Routine measurement takes time and funds away from clinical work.

These concerns were addressed in a reply by Coombs and Meehan (2005), and many have widely published counter-arguments. Coombs and Meehan pointed out the importance of clinical judgment in using routine measures and of using the opportunity

to enhance the therapeutic relationship. Indeed, the results from a measure should never be the sole basis for clinical decisions. They challenged Lakeman's claim of poor validity for the HoNOS measures and argued that the scales represented common problem areas for clients and were a compromise between specificity and feasibility. Outcome measures do not purport to capture all aspects of a client life, and indeed this is impossible. What they do provide is a "snapshot" of selected aspects, which can be compared with later snapshots to obtain a picture of how things are changing. The variety of measures available for tracking different types of outcome were highlighted by Coombs and Meehan, including those Lakeman linked to a recovery perspective. Generally, they encouraged nurses to critically reflect and learn more about routine outcome measurement in order to gauge its usefulness for their practice – with clients or in managerial roles – and the potential of measures to resolve methodology issues in some areas of mental health research. Their rebuttal was consistent with the literature mentioned previously on the benefits of outcome monitoring for multiple stakeholders.

United States of America

The situation in the USA is much more fragmented. There is no national routine protocol; some states having implemented routine measurement and others having individual agencies who decided to use measures. Several services have published their experience and, overall, they are similar to those of the UK and Australia.

Legislation mandated the use of outcome measures in publically funded mental health services in California in 1994. A later investigation into clinician attitudes towards outcome measurement highlighted the necessity for ensuring measures are clinically useful. Garland, Kruse, and Aarons (2003) found 92% of clinicians did not use the information from the mandated measures, although over half found the process of administering measures clinically useful. They noted that the main barriers to using the data were concerns about feasibility, a perception that the measures were not valid, and difficulties interpreting the meaning of the clients' scores (i.e., the measures were not user-friendly). As a result, participants in this study suggested good outcome measures should be brief, use simple language, and have clear graphic or narrative interpretation of the scores. Garland and colleagues recommended, for other services considering outcome measurement, that clinicians have input into the selection of measures and training in their use following implementation.

At variance with this are studies of client attitudes towards outcome measurement. Eisen, Leff, and Schaefer (1999) reported 90% of adult clients in Massachusetts mental health services were satisfied with the measures and many appreciated that the service was interested in how they were doing. This is consistent with an Australian study (Guthrie, McIntosh, Callaly, Trauer, & Coombs, 2008) which reported the majority of clients thought routine outcome measurement helped clinicians understand them better (76%) and led to their receiving better care (66%), and clients found discussing the results useful (60%).

The state of Virginia has worked towards outcome measurement in CAMHS since the late 1990s and found their approach to designing the protocol worked well because they involved all stakeholders (including clients and their families, providers and agencies, and higher level administration; Koch, et al., 1998). Multi-stakeholder groups were recruited to work through four steps for selecting measures (similar to the steps taken by Bickman, et al., 1998 in Australia), which were:

1. Identify what to measure – areas/domains
2. Critically review measures covering those domains/areas
3. Identify the “best” measure for use in Virginia’s public system
4. Recommend strategies to promote use across the state

Once the “best” measures in each domain/area were selected, they were piloted to evaluate their feasibility, usefulness, and cost effectiveness in CAMHS settings. The measures were “viewed positively by consumers” (p. 158) in parent focus groups, and “consumers wanted to ensure [measures] provided them with an opportunity to report on their progress, not just rely on clinician ratings” (p. 159).

The state of Ohio has an extensive routine outcome measurement programme for all public mental health services, including those for children and adolescents. The Ohio Scales were specifically developed for mandated routine use in CAMHS across the state and the data for every client is uploaded into a central databank. This is publically available online and summary reports on a wide range of outcome variables can be produced by accessing the Ohio Department of Mental Health website (<http://mentalhealth.ohio.gov>). The success of this measure has led to its use in other USA states (e.g., Texas Department of Mental Health and Mental Retardation, n.d.), and other countries, including Ireland and Canada (Healy, 2007), and has been translated into seven languages. In a survey of the state’s top performing services (i.e., those who had the highest measure completion rates), clinicians identified factors

which facilitated implementation of the protocol and those that served as barriers (Healy, 2007). The top three facilitators of measure completion were 1) management support, 2) having a reminder system, and 3) clinician buy-in. The biggest barriers were 1) consumer resistance, 2) clinician resistance, and 3) lack of user-friendliness. Healy (2007) noted that a high level of consultation with stakeholders was important to the success of the Ohio outcomes initiative and encouraged other organisations introducing routine outcome measurement to follow their lead throughout the preparation and implementation phases.

New Zealand’s experience

MH-SMART and CAOS

In 1997, the Ministry of Health released a national mental health plan that included a key objective to promote quality improvement. In order to begin advancing the principles of the plan, work was started to establish a system to collect outcome information (Kriebler, 2003). The *Mental Health Standard Measures of Assessment and Recovery Initiative* (MH-SMART) was tasked with introducing and promoting a culture supportive of using outcome measures and to provide training and resources for clinicians in the use of New Zealand’s first mandated measures (Te Pou, 2009). New Zealand followed the UK and Australia by introducing the HoNOS suite of measures, which have been mandated for national routine use here since 2005.

The large-scale trial of the introduction of the HoNOS measures was one objective of the *Classification and Outcomes Study* (CAOS; Gaines, et al., 2003). The study trialled HoNOS-CA, among others, in eight District Health Board CAMHSs across New Zealand. The authors reported there was acceptable clinician compliance for HoNOS-CA, but because this was not the case for other (non-HoNOS) measures trialled, the authors suggested further work was needed to determine whether HoNOS-CA could provide sufficient information as a stand-alone measure. Further analysis of the outcomes data from the CAOS sample was reported by Trauer, Eager, Gaines and Bower (2004). They found around 97% of “episodes of care” in community services had a HoNOS-CA completed at least once, but only around 53% had *both* start- and end-of-episode administrations. The authors said these results could be partially explained by the “snapshot” nature of the study. However, over 1300 episodes had HoNOS-CA ratings at the start and only 109 had end ratings. Even for a snapshot, this appears to be a highly unequal representation of episodes to have begun (but not

ended) and ended (but not started) during the months the study ran. This has negative implications for the utility of the measure as a progress or outcome monitoring tool, because a single administration (at either the start or end) is not useful for the clinician and client, or for use at other levels. A lack of feasibility for routine use or poor practitioner understanding of the clinical uses of HoNOS-CA could be inferred from the small number of repeat administrations. It may also suggest a detrimental impact on compliance and regular use once the study was over.

Other concerns raised by the findings of Trauer et al.'s (2004) study related to the psychometric performance of HoNOS-CA with this New Zealand sample. Confirmatory factor analysis revealed the data did not fit well with the previously proposed structure, or other possible structures that were trialled. In addition, the three subscales measuring domains of mental health had fair to poor internal consistency (Cronbach's alpha: Behaviour, .52; Impairment, .31; and Symptoms, .34). The Social subscale, which includes typical functioning domains associated with friends, family, school and self-care, was slightly better at .62, but still lower than is usually accepted, even for research purposes (i.e., .70; Groth-Marnat, 2003). The two-item Information subscale and total scale score (which does not include Information) achieved the only acceptable levels of reliability, at .81 and .76, respectively. More optimistically, HoNOS-CA scores correlated with the *Children's Global Assessment Scale* (a single item reflecting the clinician's judgment of the client's global functioning) as predicted (-.15 to -.46 for individual items, -.30 to -.55 for subscales, and -.56 for the total scale score).

The mean total start and total end scores were compared (using the subset of data where both ratings were collected) to provide an indication of HoNOS-CA's ability to track change. The authors report a significant decrease from a mean start total score of 13.2 to a mean total score of 10.9 at the end. They rightly point out that with a very large sample almost all differences would be statistically significant. This is important here as it is unclear whether a 2.3 difference would be *clinically* significant, particularly as this is less than the standard error of measurement reported for HoNOS-CA in CAMHS settings (Hanssen-Bauer, Aalen, Ruud, & Heyerdahl, 2007). No information about mean change for individual clients was reported, and no correlations with other concurrent outcome or criterion measures were investigated. Despite the authors concluding that "there are no technical reasons why any of the measures are not appropriate for routine collection" (Trauer, et al., 2004, p. 107), the question of HoNOS-

CA's reliability and sensitivity to change with New Zealand young people still seemed unanswered.

Merry and colleagues' report

A team of researchers led by Sally Merry (Merry, et al., 2004) of the *Werry Centre for Child and Adolescent Mental Health* at Auckland University investigated the use and acceptability of outcome measurement in CAMHS settings. Their important report detailed the findings from surveys, interviews, and focus groups with practitioners, parents, and young people around the country about measures being used at that time, and made recommendations for the future of routine outcome measurement in New Zealand services. This was the first time this type of information had been published about outcome measures for New Zealand CAMHS. Before outlining the findings of that report, a brief recap is warranted to reiterate the international context in which this research occurred.

The field had been reviewed in the UK (Hunter, et al., 1996) and the lack of appropriate measures led to the development of HoNOS-CA (Wing, et al., 1998). Around the same time the SDQ (Goodman, 1997) and Ohio Scales (Ogles, et al., 1999) were being developed in the UK and USA, respectively, and Bickman and colleagues (1998) were reviewing potential measures in Australia. Bickman considered all three measures (amongst others) and recommended the Ohio Scales, but the HoNOS-CA and SDQ were introduced. During this time, Australia had also finished an innovative and precedent-setting investigation into how "casemix classifications" could provide a more equitable basis for funding public mental health services. Given their success, the equivalent New Zealand project, CAOS (Gaines, et al., 2003), was modelled on the Australian methodology. The outcome measures trialled in New Zealand – as a secondary objective to the development of a casemix classification – were not chosen on the basis of independent systematic review as was done elsewhere, but simply because they were already in use in Australia and the UK (Krieble, 2003). Thus, HoNOS-CA was first introduced to New Zealand CAMHS, with mixed results (Trauer, et al., 2004). By this time, the Ohio Scales had been validated on a large scale and was in routine use in the USA (Ogles, Melendez, Davis, & Lunnen, 2001). Merry et al.'s report was released in 2004, in time to inform the national mandating of outcome measures for New Zealand mental health services in 2005.

Merry and colleagues surveyed CAMHS practitioners with the intention to obtain recommendations about measures already being used and attitudes towards outcome

measurement. The *Child Behavior Checklist* (CBCL) was the most frequently recommended measure for completion by the young person, their parents and teacher, as was HoNOS-CA for clinician completion. Low proportions of participants actually used the measures they named on a routine basis (HoNOS-CA and CBCL 16%, GAF 7%, SDQ 5%, Conners' 4%, Beck Depression Inventory and Children's Depression Inventory 0%). After identifying a measure, participants were asked to describe its advantages and disadvantages. Using the data reported by Merry et al. for each dis/advantage for each measure, the percentage of responses for each dis/advantage across all measures was calculated and is presented in Table 1.

Table 1 *Proportions of Participants* Stating Each Reason to Use/Not Use (a) Any of the Measures Identified, (b) HoNOS-CA, and (c) SDQ*

	<i>All measures</i>	<i>HoNOS-CA</i>	<i>SDQ</i>
	%	%	%
Advantages			
Specific characteristics: developmentally appropriate, positive, multiple informants, comprehensive	45.8	37.1	61.9
Ease/simplicity of administration/completion	34.1	37.1	47.6
Assists in diagnosis: screening, identifies needs, supports clinical interview	33.9	17.1	4.8
Tracks changes/outcome	32.2	31.4	-
Psychometrics: validity, reliability	19.4	17.1	19.0
Widely used	7.7	5.7	-
Promotes discussion between client and clinician	4.2	2.9	-
Inexpensive, free to use	1.2	2.9	-
Disadvantages			
Complicated, incomprehensible	23.9	25.7	4.8
Specific characteristics: too broad, too narrow	15.6	-	33.3
Intentional bias (client/clinician), subjectivity	13.2	20.0	-
Used by untrained people, no training available	11.4	14.3	4.8
Expensive	11.1	-	-
Criticisms of psychometrics, cultural bias	9.5	20.0	4.8
Misleading, out of context	8.8	2.9	-
Lack of sensitivity to change	1.7	-	4.8
Unavailable	1.1	-	-

* data taken from Merry et al. (2004).

These results suggested the top two reasons New Zealand practitioners decide whether to use or not use a measure were ease of use and its specific characteristics. In relation to the measures used in Australia and UK CAMHS, more participants saw these as advantages for using the SDQ than HoNOS-CA. In addition, the main reason

for not using the SDQ was its breadth, whereas HoNOS-CA was seen as being complicated and subject to bias. Additional findings in their study were that 97% of participants felt outcome measurement in CAMHS was “somewhat important” to “very important”, mostly because it supported service delivery and clinical work. This was extremely encouraging, however around 70% of participants were concerned the data would be misused.

In focus groups with young people and parents, the CBCL was seen as very comprehensive, but too long and difficult to complete. The HoNOS-CA was praised for its breadth and ability to track change, but concerns about the accuracy of clinician ratings, completion time, and double-barrelled questions were raised. The authors reported “wide acceptance” (Merry, et al., 2004, p. 79) of the SDQ as it was brief and simple, yet covered important issues and included strengths.

The drawback to the methodology of the study was the risk that potentially useful measures were not considered by the research team because they were not known about or identified by participants. This might explain the absence of the Ohio Scales in the results because although the measure was known to the researchers at that time, it is likely to have been unfamiliar to practitioners here.

Merry and colleagues concluded there was positive support for the notion of outcome measurement among New Zealand CAMHS practitioners, young people and their parents. Like previous researchers, they found no measure was ideal, but supported the mandating of HoNOS-CA and the SDQ for routine use in CAMHS. They noted that neither measure had been validated for use in New Zealand and recommended this be done.

Implementation of routine outcome measurement in New Zealand

New Zealand mental health policy has a commendable emphasis on valuing client perspectives and clients are noted as the principle stakeholders in mental health information systems (Ministry of Health, 2005). This is consistent with the consumer advocacy movement and the principles of the Treaty of Waitangi. In terms of outcome measurement, engaging clients in the design and implementation of a routine protocol is integral (Ministry of Health, 2005). To date, the measures introduced for routine use have been clinician-rated and there are projects underway to develop consumer-rated measures (Gordon, et al., 2004; Niumata-Faleafa & Lui, 2005), although these have been for use in adult services. HoNOS-CA has had a client-rated version developed in

recent years (Gowers, Levine, Bailey-Rogers, Shore, & Burhouse, 2002), but this has not been introduced to practice in New Zealand.

In addition, the implementation of measures for routine use in New Zealand has included a focus on ensuring clinicians find them useful in their work. The first, and current *National Mental Health Information Strategy* has a strong emphasis on information being used, not just collected (Ministry of Health, 2005). The strategy explicitly acknowledges that data must be useful for clinicians and clients in their day-to-day work together if it is to be collected in a quality manner and therefore to also be of use for management and policy level decision-making. Indeed, the primacy of this “frontline” data is clear in the Strategy: “...the quantity of information needed at regional and national levels should be a *by-product* of the information collected locally” (emphasis added; Ministry of Health, 2005, p. 3). It is laudable that this emphasis has also come through to routine practice. For example, Counties-Manukau DHB psychologists have worked to develop a model for multidisciplinary teams to make better use of the data obtained with HoNOS-CA (Stewart, 2007). This is an excellent example of routine outcome measures informing decision-making at all levels of the mental health system.

However, as in other countries, routine outcome measurement with HoNOS-CA has not been met with universal support by clinicians in New Zealand. Egelston and Watkins (2008) outlined a number of criticisms of the mandated measurement protocol, including the manner in which HoNOS-CA was selected and implemented. They were disparaging of the methodology of Merry and colleagues’ (2004) and cast doubt on the appropriateness of the conclusions drawn in the report. In addition, Egelston and Watkins raised concerns about HoNOS-CA’s psychometric properties with New Zealand young people, its sensitivity to change, clinical utility, adolescent-oriented item content, and acceptability to clinicians. They concluded that the implementation of routine outcome measures in New Zealand CAMHS was premature and that HoNOS-CA had significant limitations and therefore other measures should be considered.

Introduction to the present research

Routine outcome measurement in mental health services has become part of standard practice internationally in response to pressure from multiple sources for evidence that services are both efficient and effective. New Zealand is no exception and has

followed trends elsewhere in introducing HoNOS-CA for use with all CAMHS clients since 2005. National policy here is consistent with literature suggesting client perspectives should be sought and that measures should have utility for clinicians, management, and funders, and be culturally appropriate.

However, the measure selected for routine use in CAMHS here can be critiqued for not meeting the aspirations of New Zealand mental health policy or the requirements of a good outcome measure. HoNOS-CA has mixed psychometric support in the international literature and CAOS revealed poor reliability and weak factor structure when used with New Zealand young people. In addition, others have criticised HoNOS-CA for its clinician-rated design, subjectivity, administration difficulty, and lack of utility at a clinical level for treatment planning and monitoring. These are all crucial features to consider when selecting an outcome measure (Koch, et al., 1998) and areas where HoNOS-CA presently falls short. Furthermore, HoNOS-CA was not selected based on any consideration of the literature as happened in other countries. Consequently, other potential measures may have been overlooked or may have been developed in the years since HoNOS-CA was selected.

This situation highlights gaps in the outcome measurement field for New Zealand. The present research aimed to remedy this by examining the measures currently available for use with children and adolescents that are suitable for routine use by multidisciplinary teams. This search and review of measures was intended to seek those meeting the requirements of a good outcome measure. That is, a measure that meets the needs of funders, management, clinicians and clients by being psychometrically sound, useful, and feasible for routine use. This was particularly important, as no systematic search for measures had been conducted for New Zealand services. More generally, the research aimed to provide this updated overview of the available measures because similar previous research was last published over a decade ago (i.e. Bickman, et al., 1998). Most measures have been developed overseas so there can be difficulty transferring a measure “as is” to our local context. Therefore, in keeping with good practice, preliminary investigation of the selected measure aimed to ascertain its degree of appropriateness for use with a New Zealand population.

Based on a series of six studies, the ultimate aim of this research was to provide evidence-based recommendations on the outcome measure most suited for routine use – or adaptation for routine use – with young people in New Zealand. As has been

found in previous projects, it was predicted that no measure would be ideal for immediate use here. However, the results of the present research were expected to suggest at least one potential measure and indicate what further work would be required. Therefore, to summarise, the research had three aims (with reference to the relevant studies added below):

1. Provide an updated overview of potential routine outcome measures available for use with children and young people.
 - Chapter 6 – Study Three: The search for suitable measures (p. 95)

2. Select at least one that fulfils the requirements of a good routine outcome measure (i.e., psychometric strength, usefulness, and feasibility).
 - Chapter 4 – Study One: Current HCN measurement practice (p. 42)
 - Chapter 5 – Study Two: Current HCN measurement practice (p. 72)
 - Chapter 7 – Study Four: Consultation and selection of an outcome measure (p. 125)

3. Obtain preliminary information about the measure's appropriateness for use with a New Zealand population.
 - Chapter 9 – Study Five: Preliminary local data for the Ohio Scales (p. 149)
 - Chapter 10 – Study Six: Informant feedback on the appropriateness of the Ohio Scales for New Zealand (p. 160)

CHAPTER 3 – OUTCOME MEASUREMENT IN MULTIDISCIPLINARY SETTINGS

The next four chapters describe a series of studies undertaken for a government agency, the *High and Complex Needs Unit* (HCN), based in Wellington. These studies concerned the use of outcome measures with teams of their practitioners from various disciplines and the search for a measure that would be useful for work with young people in this context. The outcome measurement needs of these multidisciplinary teams and what ideal characteristics they sought in a measure are described shortly. Then, the HCN, their work, and the tender this project was based upon are outlined. A brief overview of the HCN studies concludes this chapter and sets the scene for those that follow.

Outcome measures for young people and multidisciplinary teams

To be suitable for routine use, outcome measures should meet certain criteria. As described previously, these criteria include psychometric strength, usefulness, and feasibility. Measures for young people have additional special requirements. Given the systems within which children and adolescents live, teachers, therapists, parents and/or caregivers often have a valuable contribution to make to the tracking of outcome. Although outcome measures can usually only be completed by older

children or adolescents, therapeutically, it is important to at least talk with all children about their view of how the intervention is working. Obtaining multiple perspectives can make outcome measurement a more time-consuming process, but the similarities and differences in the results can provide an opportunity for discussion and further refinements to an intervention (Trauer, 2003). Outcome measures for young people also need to be developmentally appropriate (Ollendick, Grills, & King, 2001). Unlike measures designed for adults, it is imperative to consider the needs and abilities of different aged children. This means examining the reading age of self-report forms, normative data about what level of an issue is problematic at different ages, and age-appropriate concepts and examples in the items. Ideally, measures for young people use multiple perspectives, and developmentally appropriate items and scoring procedures.

For multidisciplinary teams there are also extra issues to consider when it comes to selecting outcome measures. Psychologists have extensive training in the use and interpretation of complex psychometric instruments and therefore have access to the full gamut of outcome measures available. One option is to place the responsibility for outcome measurement on psychologists; however, a multidisciplinary team might have one or no psychologist so this is unlikely to be the most feasible option. The alternative is to use a measure that does not require advanced training, is straightforward to administer, score and interpret, and is comprehensible to all practitioners on a team, regardless of professional discipline and training. This means the burden of outcome measurement can be shared and the results of a particular administration communicated easily among team members. In addition, any measure selected should be maximally efficient as practitioners working with young people often have scarce time, funding, and other resources.

The High and Complex Needs Strategy

The special challenges involved in meeting the needs of young people with multiple and complex social, educational and health difficulties led to the development of a joint sector approach to service planning and delivery. This strategy, aptly named the *Intersectoral Strategy for Children and Young People with High and Complex Needs* ("HCN Strategy"; High and Complex Needs Intersectoral Unit, n.d.), aims to facilitate collaboration between agencies to streamline service delivery and improve outcomes. The HCN Strategy is funded by the Ministries of Health and Education, and Child Youth

and Family (CYF). Collaboration between these three sectors is promoted by providing a formal link between agencies, so they can use the funding for individualised intervention plans that are developed and implemented at a local level.

High and complex needs are broadly defined, but in practice are needs at high or severe levels that cross sector boundaries, such as having mental/physical health and educational needs, or care and protection as well as educational needs. Of course, it is the nature of children that any single need is likely to impact on multiple areas of their functioning; however, the HCN Strategy targets the extreme end of the spectrum. Often, these are children who would otherwise “fall through the cracks”. The HCN Strategy addresses the problem of a young person receiving a number of unrelated interventions from each sector, which might not be an efficient use of resources or result in the best possible outcomes. Many young people funded by the HCN Strategy have severe behavioural and emotional difficulties, traumatic or unsafe family histories, and some have developmental or intellectual disabilities, chronic medical conditions, or learning difficulties. Expulsion and exclusion from school is common, as are significant family/whānau needs, difficult social circumstances, economic disadvantage, and for some, traumatic or highly disruptive experiences, such as being a refugee (HCN Policy Advisor, personal communication, 2006). Many of these young people experience a combination of these issues.

The High & Complex Needs Unit

The High and Complex Needs Unit (referred to hereafter as “HCN”) was established in 2001 to administer the funding made available by the HCN Strategy. The purpose of this funding is to provide extra resources and support over and above what is normally available through the three sectors’ usual services. To be eligible for funding the young person (up to 20 years old) must have needs at a level that no single agency can meet within their usual service specifications. At the commencement of the present research, around 200 young people had received funding; between 60 and 90 at any one time (High and Complex Needs Intersectoral Unit, 2005).

With consideration of, (a) the sectors involved (b) the young person’s specific needs and priorities, and (c) their age and developmental level, an interagency team works with the help of a regional HCN Advisor to develop an individualised intervention plan. Generally, these HCN teams are comprised of the practitioners involved with the young person at the time of the funding application. It is seen as a method of formalising the

collaboration of practitioners who already have input into the situation. Once eligibility has been determined and an intervention approved by HCN, funding is usually offered for 12 months. The intention of HCN is for this funding to allow intensive collaborative input to reduce the young person's needs to a level that the usual services in each sector can address; thus, it is not seen as long-term input. While in theory this is an admirable goal, in reality, many of these young people require a second or even third year of funding (HCN Professional Practice Advisor, personal communication, 2007).

HCN's request for tenders

HCN staff reported that anecdotal evidence showed the interventions were generally helping manage or improve young people's situations. However, they had no objective, standardised, or routinely collected data to support this. Interventions involving intersectoral cooperation and providing function-specific services rather than agency-specific services have been effective overseas (for example, between Health and Education, see Young, Earle, & Dadebo, 2004). However, in New Zealand, evidence for the effectiveness of this approach was less clear. Preliminary work by Calvert and Lightfoot (2002) suggested a positive impact for young people with high and complex needs, although the general effectiveness of HCN and how individual interventions could be improved, was largely unknown.

In addition, HCN interventions involve considerable amounts of planning and professional time so are both labour intensive and costly. Not having outcome data made a cost-benefit evaluation difficult for the HCN staff who are charged with administering limited funds, time, and people efficiently. At a more individual level, HCN teams were missing vital feedback about the success of their intervention, and the opportunity to change direction if necessary, by not having access to outcome data.

Being unable to measure outcomes meant HCN was unable to evaluate fully the contribution of teams' interventions to the vision of the HCN Strategy. HCN recognised these multiple opportunities for using outcome data within their organisation and disseminated a request for tenders to provide a report recommending measures for HCN teams to use. The tender document outlined the aims and requirements of HCN with regard to outcome measurement (High and Complex Needs Intersectoral Unit, 2005). While they acknowledged the difficulty of finding a single outcome measure to suit all young people, they asked measures be sought that represent the "best possible

fit” (p. 5) to the well-being framework of the HCN Strategy. The well-being framework is comprised of seven domains,

- Physical health and mental wellbeing
- Attachment and belonging
- Cultural identity
- Development and independence
- Social interaction
- Strong educational foundations for lifelong learning
- Safety and security

They sought measures for use at both the individual progress monitoring level and at the aggregate level, so outcomes could be analysed across the HCN population and sub-groups (age, ethnicity, or gender). At the individual level, the intention was for measures to cross sector boundaries and be useful for informing reviews of intervention effectiveness. At the aggregate level, the intention was to describe the overall impact of funding, be useful for policy and evaluation decisions, and to indicate the direction and extent of outcome trends.

As well as clearly describing the purpose and potential use of measures, HCN specified a number of considerations for measures being identified. These were adapted from those in a previous CYF internal report on outcome measurement (Barber, 1998, cited in High and Complex Needs Intersectoral Unit, 2005). In brief, these considerations were:

- Feasibility
- Accommodation of missing data
- Interpretability
- Quantifiability
- Repeated unscheduled administration
- Appropriate norms
- Is “culture fair”
- Reliability and validity

Although searching for and recommending suitable outcome measures for use in such a diverse service appeared complex, a series of studies were designed to do this and the proposal won the tender.

The HCN Outcome Measurement Project

The proposal outlined a series of studies that, together, would give enough information to make initial recommendations to HCN. Information about what was currently happening in teams, in terms of outcome measures, was thought to be a crucial starting point. HCN gave examples of commonly reported outcomes such as, “seems much more settled”, or “is doing better at school”. This information highlighted the necessity of speaking with practitioners about what they understand about monitoring progress and what they use to indicate outcome. In the final proposal, focus groups with both the regional HCN Advisors and practitioners around the country and a postal survey of all past and present HCN team members (practitioners) were planned. This was intended to provide information about current attitudes to the use of measures, and to help identify which measures were currently being used and were recommended by practitioners.

To identify suitable outcome measures an extensive search of the international literature was proposed. The intention was to compile a master list of measures for use with young people then narrow it down according to the requirements of a good routine outcome measure. With a shortlist of possible measures, it was proposed to combine feedback from the survey and focus groups with expert review in order to make a final selection.

The project was begun in mid-2006, with ethical approvals obtained and all data collected during that year. The project went smoothly and after feedback on the draft report from HCN and two national peer reviewers, the final report was presented in June 2007 and made publically available via the HCN website in 2008. The next four chapters are adapted from the final report.

Introduction to the HCN studies

As mentioned above, the series of studies that fulfilled the tender requirements are outlined in the following chapters. As each had distinct individual goals, methodology and results, they are presented separately. *Chapter 4* outlines data gathered in Study One using a postal survey about the current outcome measurement practices of HCN practitioners. Practitioners also gave valuable suggestions for the implementation and use of measures with young people and many of the comments were consistent with issues raised in the literature. *Chapter 5* is concerned with Study Two: five focus

groups conducted around New Zealand with HCN practitioners to gather more specific feedback about how they view outcome measurement and how they currently evaluate change with their clients. *Chapter 6* describes Study Three, which was the lengthy process of identifying potentially useful measures and reducing this to a manageable short-list for further review. Study Four is covered in *Chapter 7* where the process of consultation with a group of experts and HCN practitioners about the short-listed measures is explained. In addition, the chapter concludes with the selection of the outcome measure best suited for use with these multidisciplinary teams, based on the results of all the HCN studies.

CHAPTER 4 – STUDY ONE: CURRENT HCN MEASUREMENT PRACTICE

HCN recognised how important being able to reliably detect changes occurring during interventions with young people is for sound decision-making and planning and sought an outcome measure to do this. Like for other collaborative, multidisciplinary services for young people with multiple needs, the search for an appropriate measure was complicated. Others have tackled this issue and often the starting point is to elicit the views of frontline practitioners. As Bickman and colleagues (1998) reported, the effective use of measures depends upon practitioners feeling it is a worthwhile activity to engage in. It is important for stakeholders at all levels to be prepared for and committed to the implementation of a new monitoring protocol (McDaniel, 1996; Rock, et al., 2001). Therefore, the opinion of HCN practitioners was sought in the present study as they were seen as the best source of information about measures currently being used, and about issues to consider in the choice of an appropriate measure.

In their project, Bickman and colleagues (1998) talked with stakeholder groups around Australia. Most stakeholders were supportive of outcome measurement in mental health services for young people, but expected frontline practitioners to be the source of strongest resistance. The reasons thought to underlie this resistance included anti-empiricism, fear of change, paranoia about what the data would be used for, lack of measurement expertise, taking time away from clinical practice, and lack of relevance for their client work. However, clinicians themselves were actually generally supportive

of routine monitoring, but felt certain conditions must be met before they would use measures in their own practice. Clinicians felt measures needed to cover symptoms, functional impairment, and family functioning. They were also concerned about the time and administrative burden, the training required, ensuring robust psychometrics, and the need for simple measures that are easy to use and interpret. Some felt outcome measurement could negatively affect the therapeutic relationship. A minority thought it was not possible for a single standard measure to be appropriate for all clients, and that measures were unable to “capture the full picture” (Bickman, et al., 1998, p. 57) of individual client’s lives.

Later, Merry and colleagues’ (2004) project here in New Zealand surveyed psychologists, social workers, and counsellors about the measures they currently use. The purpose was similar to Bickman et al.; that is, to recommend measures suitable for routine nationwide use in CAMHS. Although Merry et al. asked what participants felt were the advantages and disadvantages of the specific measures they named, opinions on the notion of routine measurement was not sought. It is possible that practitioners may have identified strengths and weaknesses for particular measures but still not felt they were a good routine outcome measure. The Merry et al. project provided an important beginning to New Zealand research into routine outcome measurement for child and adolescent services, which needed to be built upon.

Another New Zealand project surveyed psychologists, counsellors, and psychiatrists about their opinions and use of measures in therapy for the ACC Sensitive Claims Unit³. Hodgetts and colleagues’ (2005) postal survey asked participants to rank order the importance of a number of sources of information about client progress, and provide comment on outcome measures and measurement in general. They found that clinicians ranked subjective indicators of progress in therapy as most important, with self-report questionnaires only having a modal ranking of 11th (of 14 indicators). Encouragingly, 58% of participants said they used measures in their practice and a very diverse range of formal and informal monitoring approaches were reported by participants. There was a degree of concern about the notion of regular outcome measure use, with clinicians citing inadequate psychometric properties, lack of suitability for New Zealand clients, lack of training or access to measures, and the role of ACC in the process, as potential barriers to be overcome. The authors noted that

³ A subsection of the Accident Compensation Corporation who funds rehabilitation (including counselling) for young people and adults who have experienced “mental injury” as a result of sexual abuse/assault.

participants appeared to “misunderstand what monitoring outcome means” (Hodgetts, et al., 2005, p. 38), as many responses were not consistent with the outcomes or assessment literature. Despite this, participants indicated support for the introduction of routine measurement, contingent on the measures being suitable, feasible, useful, and available to clinicians. The results of this survey were used to help guide the later search for, review, and recommendation of measures in the same larger program of research (i.e., Fitzgerald, et al., 2006).

The present study aimed to identify practitioners’ opinions about the usefulness and feasibility of measures, and the potential barriers and benefits to using them, using a postal survey. Information was also sought about how they currently monitor client progress and make decisions about ending the intervention. A mix of open-ended and Likert-type rating scales were used in order to build upon the work done in New Zealand and Australia previously. Given the findings outlined above, it was expected that participants would express conditional support for routine measurement and that a minority would express resistance to the notion. It was hoped that a number of measures would be named and critiqued by participants, with their recommendations then able to inform the search and short-listing of measures conducted subsequently.

Method

Participants

HCN provided two sources of practitioner names that comprised the group from which survey participants were recruited. The first was a list of 465 service providers who had worked on interagency teams since HCN began (approximately five years previously). Thirty-nine parents, caregivers, and non-professional service providers were excluded, which gave a total of 426 potential individual survey participants. The second source was a list of contact details for 36 Local Service Coordinators (LSCs) – the people responsible for managing the local operations and interventions for the 64 (then) current HCN teams. Information regarding the professional disciplines or other demographic details of the potential participants was not available.

Packs including an information sheet and survey form were sent to all individual practitioners. In addition, each LSC was sent a bundle of six packs to distribute to practitioners in the HCN team they coordinated. Some LSCs received more than one

bundle as they coordinated more than one team (12 LSCs received two bundles, one received three bundles, two received four bundles, and two received five bundles). Of those sent out, 25 individual packs and three LSC bundles were returned due to an incorrect address. Twelve practitioners made contact to request a survey pack.

Seventy-two completed surveys were received from practitioners, a return rate of approximately 18%. However, it is difficult to determine the return rate accurately as data were not available regarding how many surveys were not received by the practitioner (for example, due to an out-of-date address) and how many surveys were not distributed by LSCs (for example, because their team members had already received a pack in the individual mail out). Therefore, this return rate is likely to represent an underestimate.

Materials

Participants completed the *Survey on Outcome Measurement Use* (see Appendix A), which was developed for the purposes of this study. There were three steps in the development of the survey. First, an initial draft was created with regard to the needs of HCN. Second, this was sent for critique to two expert advisors and practitioners representing education, health, and welfare organisations. Third, the final draft was sent to HCN Advisors and representatives from government sectors involved in HCN services for their review. Suggestions for improvement were incorporated into the final survey.

The survey contained 20 questions covering the characteristics of the participant, characteristics of their HCN clients, and their views on outcome measurement. Short-answer questions gathered information about participant characteristics such as practice discipline, qualifications, experience with HCN, geographic location, and type of employing agency. Further questions focussed on the participants' HCN clients, including age group, gender, ethnicity, and nature of difficulties. Ten questions concerned outcome measurement; some of these were open-ended and some required Likert-type scale ratings. Initial items covered participants' opinion about the groundwork required to use measures with HCN clients, how they gauge progress, and how they decide when their client is ready to transition to mainstream services (i.e., finish HCN-funded interventions). Participants were also asked about measures they used or knew of, to rate the importance of potential barriers and potential reasons for using measures, to indicate which domains of functioning measures could be useful

for, and to outline their ideal outcome measurement protocol. At the end of the survey was an optional section containing three open-ended questions for general feedback about outcome measurement with HCN clients and the study as a whole.

Procedure

A pack containing an Information Sheet (see Appendix A) and the survey was mailed to potential participants (or distributed through LSCs). All LSCs were phoned to ensure the survey bundles had been received and the instructions were understood. Participants were invited to complete the survey individually and anonymously. After three weeks, a reminder was posted to individual practitioners. When completed surveys were returned, they were numbered and organised for analysis. Three types of data were extracted: participant and client characteristics, and quantitative and narrative answers to the outcome measurement questions. Participant and client characteristics were summarised. Quantitative data were analysed using simple statistics (the sample size precluded more advanced analyses) and the narrative responses were organised according to content themes for each question.

Results

Characteristics of participants and their clients

Participant characteristics

The 72 practitioners who participated in this survey reported their professional discipline, ethnicity, qualifications, experience with HCN, experience in their current role, where they were employed and by which sector/organisation. Of the 69 participants who answered the ethnicity question, 46 identified as Pakeha/NZ European, seven as Māori, five as New Zealanders, three as English/UK, two as Indian, two as North American, two as Pasifika, and two from Europe. Fifteen professional disciplines were represented; the largest being psychologists, followed by teachers, nurses, social workers, and service directors (see Table 2).

Table 2 Professional Discipline of Survey Participants

<i>Discipline</i>	<i>Number of Participants</i>
Psychologist	18
Teacher	11
Nurse	9
Social Worker	9
Service Director/Manager	8
Special Education Advisor	4
Counsellor, HCN LSC, Paediatrician, Police Officer, Psychiatrist, Psychotherapist, RTLB, Support Coordinator, & Support Worker	≤2 each
Total	72

Participants had a wide range of experience working in their field, from 1-45 years ($M = 13.95$, $SD = 11.53$), and of the 66 participants who reported their qualifications, most were tertiary educated, with 36% at the Bachelor's level and 56% at the post-graduate level. Experience with HCN clients was generally limited. Forty percent of participants reported having worked with only one HCN client (range 1-28, $M = 2.89$, $SD = 3.53$), and overall had between three months and six years experience working with HCN ($M = 2.28$ years, $SD = 1.43$).

Seventy-eight percent of participants worked in the North Island. Auckland was the most frequently reported work location (18%) with Nelson/Blenheim next (13%). Ten percent of participants were in each of Bay of Plenty, Whangarei/Northland, Hawke's Bay, and Manawatu/Horowhenua. The remaining participants worked in Waikato, Gisborne, Taranaki, Wellington, Christchurch and Dunedin, giving a wide distribution across the country. Education was the best represented work setting (40%), followed by DHBs (27%), private/community agencies (19%), CYF (12%), and the Police (1%). One participant did not report their work setting, and three participants identified two settings.

Client characteristics

Participants indicated the gender, age group, ethnicity and type of presenting issue of their HCN clients. Details for approximately 200 clients were obtained. The clients were predominantly male, just over half were Pakeha/NZ European, and the majority

were aged between 11 and 15 years (see Table 3).

Table 3 *Characteristics of Participants' HCN Clients*

<i>Client Characteristic</i>	<i>Proportion of Total</i>
Gender	
Male	71%
Female	29%
Age Group	
0 – 10 years	22%
11 – 15 years	52%
16 – 21 years	26%
Ethnicity	
Pakeha/NZ European	56%
Māori	41%
Pasifika	2%
Other ethnicity	1%

In addition, participants rated how significant a range of possible presenting issues were for the HCN clients they had worked with (see Figure 2). The most significant issues reported were family dysfunction, anti-social behaviour/aggression, and learning/achievement difficulties. Areas rated most often as not significant or very low significance included physical disability/health, lack of cultural identity, internalising, communication/speech issues, and brain injury/cognitive difficulties.

The “Other” presenting issues that were mentioned were generally rated as very significant. Those not already covered by the categories provided were substance use and sexual offending (both mentioned twice) and teenaged pregnancy (mentioned once).

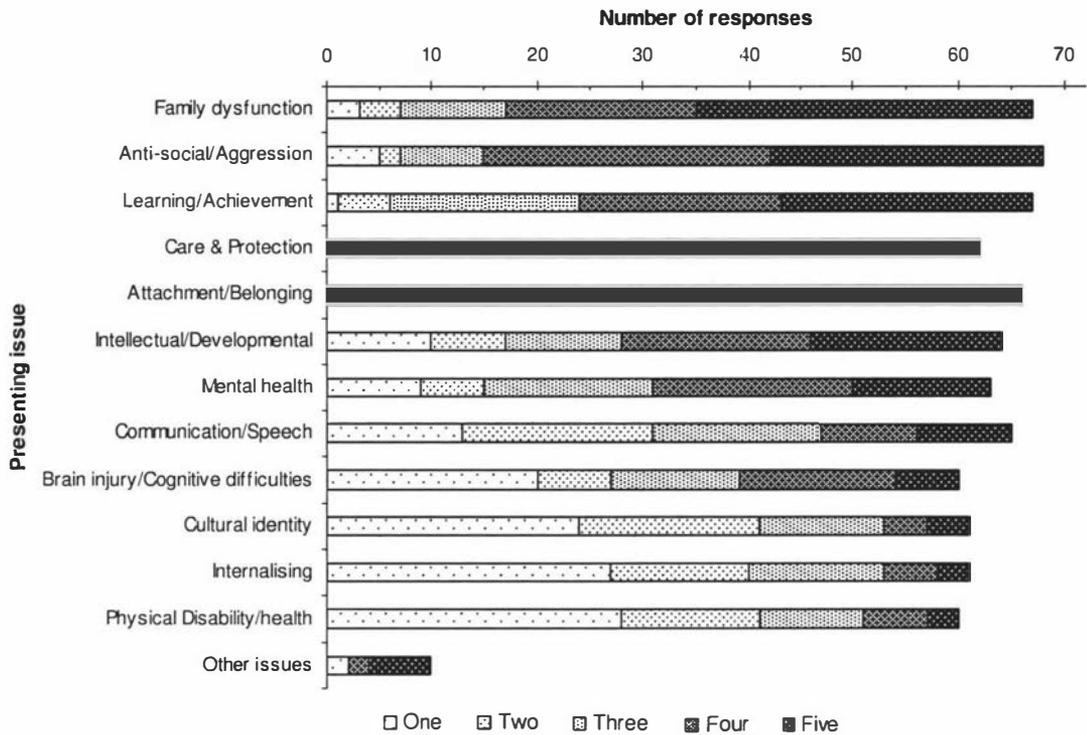


Figure 2. Number of participants rating the significance of each presenting issue for the HCN clients they have worked with. Rated on a scale from 1 (not significant at all) to 5 (very significant issue), and ranked top to bottom by mean rating.

Participants' current outcome measurement practice

Information about participants' current outcome measurement practice was elicited by asking them to describe their experience with measures in HCN interagency teams, and to indicate how they know a client is making progress and decide when to transition back to mainstream services.

Experience using outcome measures for HCN

Sixty-six participants described their experience using outcome measures with the HCN clients they had worked with. Both positive and negative experiences were recounted.

Many reported little or no experience with measures at all and the term "outcome measure" appeared frequently misunderstood. Assessment/diagnostic tools, for example the *Adaptive Behavior Assessment Scale* (ABAS), the CBCL, Conners', and *Wechsler Intelligence Scale for Children* (WISC), were all mentioned by several participants as examples of measures they use for monitoring outcome. However, a

small number identified outcome measures they use regularly, such as the Beck Inventories and HoNOS-CA.

In order for outcome measurement to be useful, the importance of building rapport with the client and explaining the reason for using measures was emphasised in a number of responses. Those who thought measures had been useful liked being able to track progress with their clients. Others used measures because it was a requirement of the organisation they worked for, but this was problematic for some, for example,

...we have difficulties in assessing level of need or ongoing risk from HoNOS-CA in particular.

Participants who thought measures were not useful commented that they were “waffley, long-winded, vague”, and felt they were too subjective. Some were concerned measures were too inflexible and narrow, which led to misleading conclusions being drawn about the client or intervention. In addition, the issue of sensitivity to change was raised and whether measures tracked important content areas, as illustrated by this comment,

The biggest thing [outcome measures] sometimes miss is the small movements that clients and families make. Things you cannot measure by ticking boxes.

There was a theme about measures not being easy to use with some clients, particularly because of their presenting issues or low reading ability. In addition, other practical reasons for not using measures were mentioned, for example,

The amount of time it takes to administer is the most significant indicator of whether or not frontline staff are prepared to use [measures]...

Many reported positive experiences as a consumer of outcome data that had been collected by an expert or specialist, often a psychologist. Conversely, some “experts” expressed frustration about explaining to other practitioners why measures should be used and objected to being placed in the role of sole “outcome measurer” for the team.

Finally, the importance of integrating data from outcome measures with other sources of information was emphasised. Many participants talked about measuring outcome using behavioural incident recording, by discussion among interagency team members or review meetings with parents and teachers, and by recording goal achievement.

Indicators to practitioners of client progress

Participants rated the importance of thirteen possible sources of information for determining whether clients are making progress. Twenty-two participants gave no rating for one or more of these items and it was unclear what the reason for this was. In addition, five participants did not answer this question (see Figure 3).

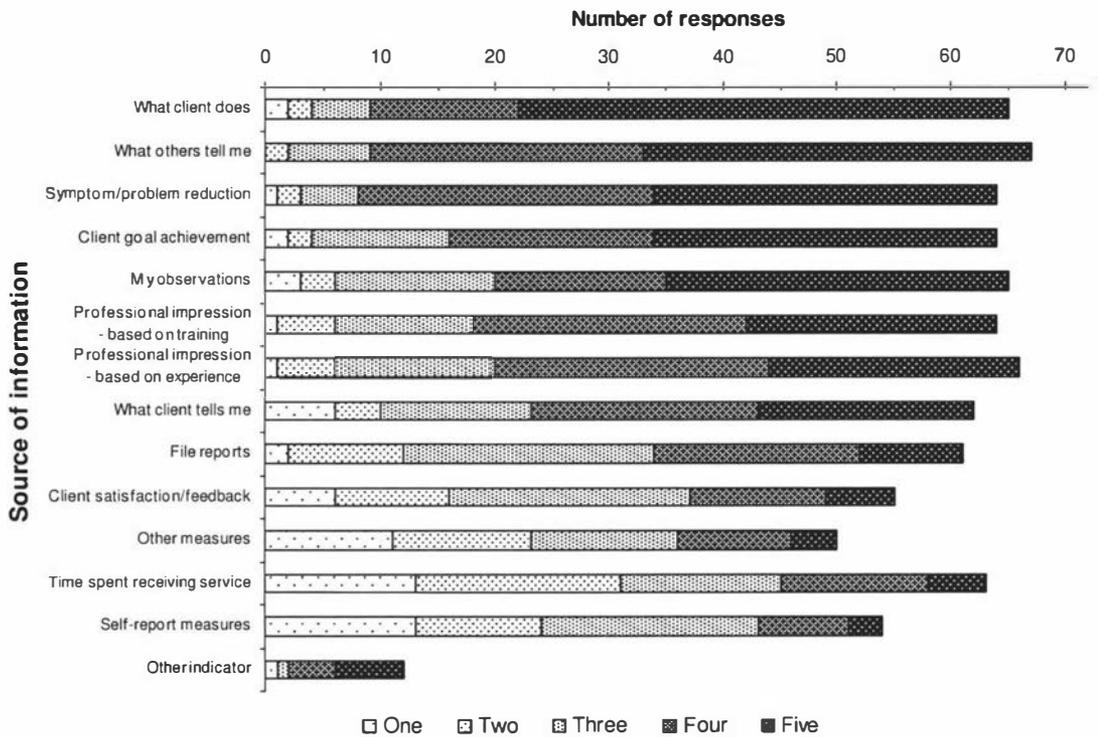


Figure 3. Number of participants rating the importance of each source of information for determining the progress of HCN clients they have worked with. Rated on a scale from 1 (not important at all) to 5 (very important indicator), and ranked top to bottom by mean rating.

Noticing changes in client behaviour and hearing from other adults involved with the client were the most important sources of information for determining client progress. Symptom/problem reduction, achieving previously set goals, and the practitioner’s own observations/impressions were important too. In general, participants did not rate self-report or other monitoring measures or the amount of service the client had received as important for indicating progress.

Thirteen participants reported “Other” indicators. One was a rating without further description and six were covered by the categories provided (for example, “parent report”). Four responses were related to changes in the systems around the client:

whānau behaviour, reduced multiagency input, stable living situation, and timely access to specialist input. The remaining responses were: an indication of strength/improvement, increased prosocial behaviour, and objective data, such as number of contacts with police.

Indicators to practitioners of client readiness to transition

Next, was a similar question regarding the sources of information they used to decide if clients are ready to transition back to mainstream services. HCN funding provides intensive interagency input to address client needs that cannot be met within the scope of usual services. As a result, it is intended to cease when the needs can again be met by these services. This survey question asked about readiness to transition, which is essentially the decision to finish HCN funding/intervention. Participants rated the same sources of information as in the progress question, but according to how important they were for making the decision to transition (see Figure 4).

As for the question about progress, client goal achievement, reduction in symptoms/problems, changes in client behaviour, and information from other adults involved with the client were rated as most important. For both questions, the ratings resulted in a similar ranking by importance for the various information sources. Again, participants did not rate self-report or other monitoring measures as important for making the decision to cease HCN funding/intervention.

Seven participants reported “Other” indicators, however five fit into the categories provided (for example, the name of a specific measure). As with the previous question, the remaining indicators were related to changes in the systems around the client: reduced multiagency input, stable living situation, and availability/suitability of mainstream services.

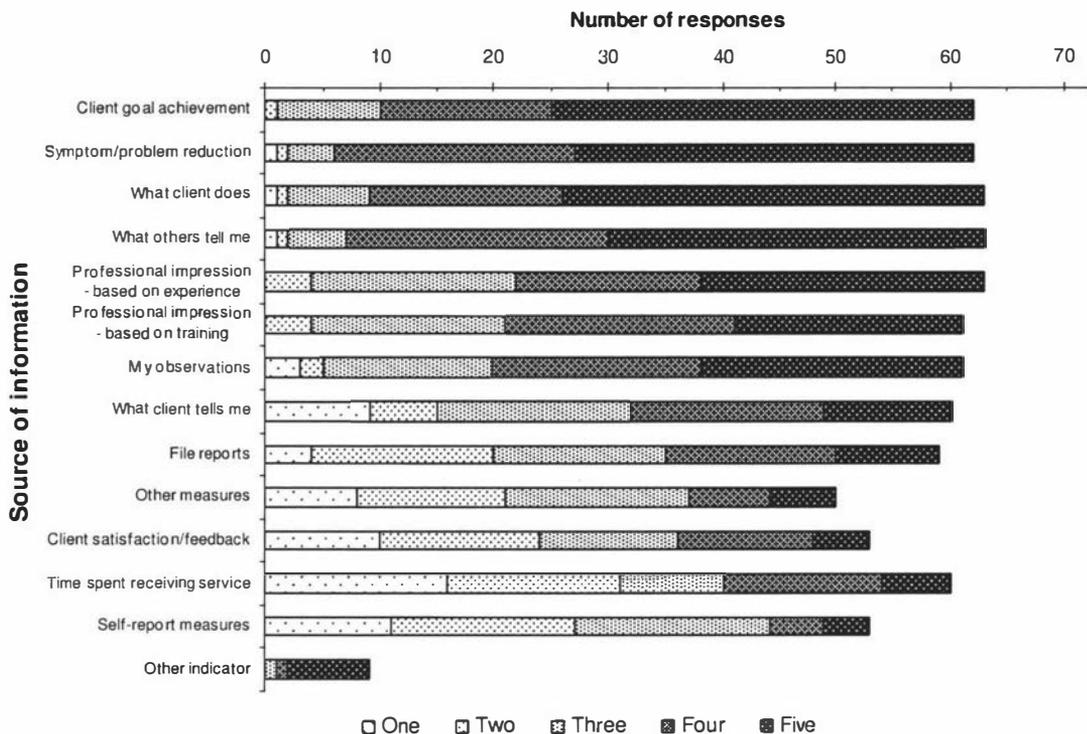


Figure 4. Number of participants rating the importance of each source of information for determining whether HCN clients they have worked with are ready to transition back to mainstream services. Rated on a scale from 1 (not important at all) to 5 (very important indicator), and ranked top to bottom by mean rating.

Reasons for using outcome measures

Participants rated the importance of ten reasons why practitioners should use outcome measures. Given the low importance assigned to measures as a source of information about progress, the results from this question were surprisingly positive. The modal rating was 5, the median 4, with a mean rating of 3.89, an overall indication that these reasons for using measures were highly important (see Figure 5).

The most important reasons for practitioners to use outcome measures were measures having specific characteristics to suit their client and that change/outcome can be tracked. The least important reasons for using measures were to promote discussion between clients and practitioners, and because a measure was widely used or inexpensive. However, these reasons were still rated as important by a large proportion of participants.

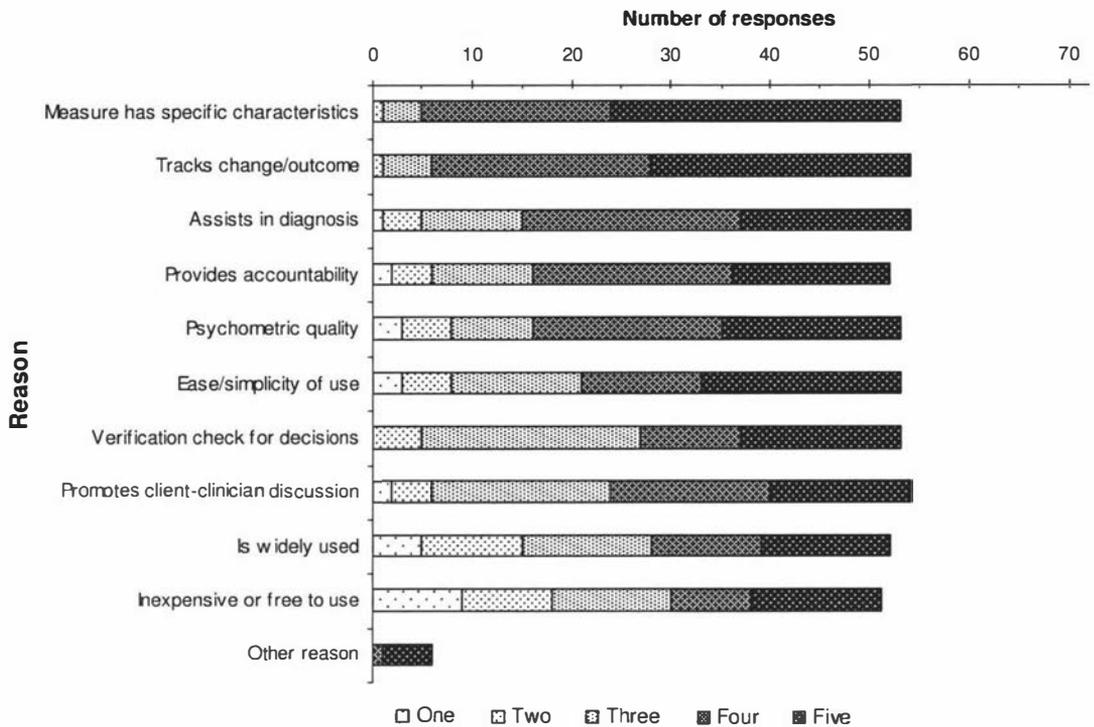


Figure 5. Number of participants rating the importance of each reason why practitioners should use outcome measures. Rated on a scale from 1 (not important at all) to 5 (very important reason), and ranked top to bottom by mean rating.

Six participants reported “Other” reasons. One response was that practitioners would use measures when they lack professional skill (i.e., skilled practitioners would not need to use measures). The remaining reasons were (1) using measures is expected because it is part of core professional training, (2) to track the therapeutic alliance, (3) to show whether goals have been reached, and (4) because it is required by funders or management.

Barriers to using outcome measures

Participants rated a list of 16 possible reasons practitioners might be prevented from using measures. The modal rating was 4, median 3, with a mean of 3.42; overall, an indication these barriers were of moderate importance (see Figure 6).

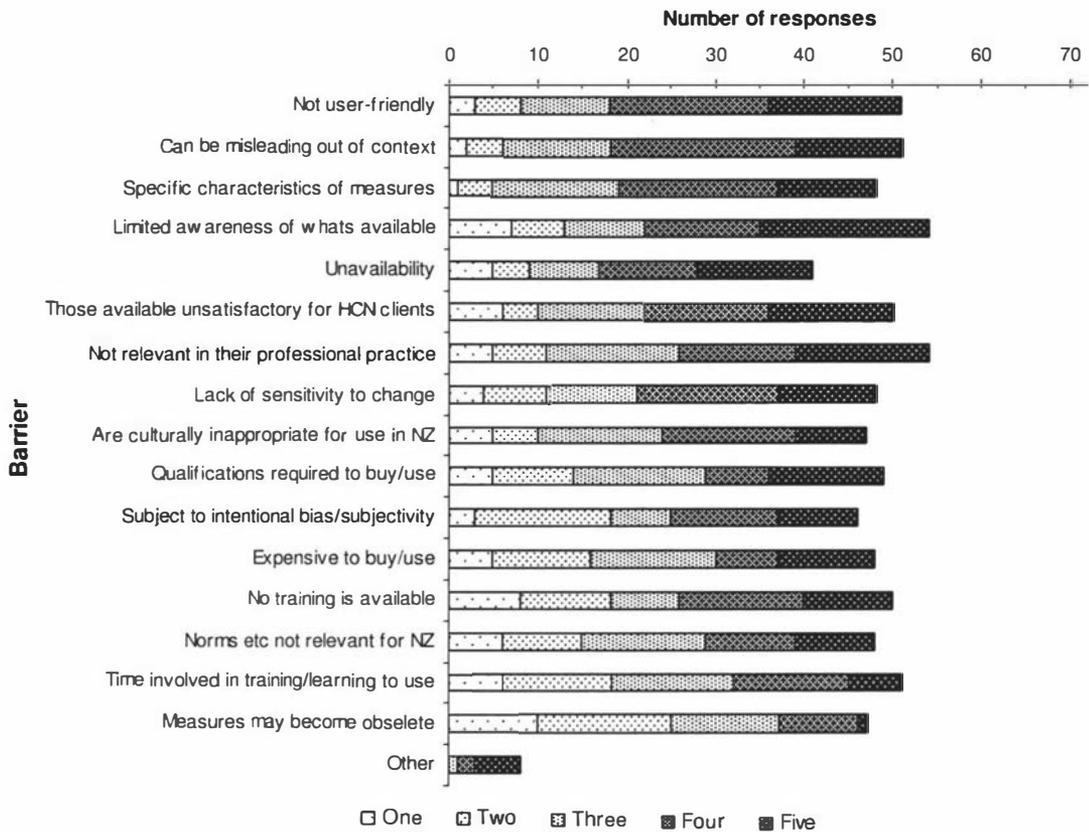


Figure 6. Number of participants rating the importance of each barrier to practitioners using outcome measures. Rated on a scale from 1 (not important at all) to 5 (very important reason), and ranked top to bottom by mean rating.

The biggest barriers to the use of measures were that they were not user-friendly and could be misleading. Other important barriers included the unavailability of appropriate measures and a lack of awareness about what was available. Spending time training to use measures and the risk that measures can become obsolete were not considered important barriers. Eleven participants suggested additional barriers,

- A single measure is not comprehensive enough
- Lack of mandate from management to use them
- Goal achievement is preferable; too hard to make problems concrete enough to measure
- Measures don't show improved functioning across settings
- Measuring outcome is not seen as standard practice
- Other professionals do it; specialists are not available to do it
- Time constraints; feasibility of using measures that take longer than a minute
- Unwillingness to do extra work

Perceived usefulness of outcome measures

In an optional section of the survey, participants considered whether a recommended set of measures would be useful for providing services to HCN clients. Their open-ended responses clearly fell into three categories, “No” ($n = 16$), “Yes, but...” ($n = 15$), and “Yes” ($n = 26$).

Of the 16 responses in the “No” category, 10 explained the reason for their position. The added administrative workload was one reason mentioned. For most however, the concern was that measures were not able to fit the individual needs of clients and were too narrow for monitoring such a diverse group. To illustrate,

No, I don't think a one size fits all approach is applicable...Very difficult to reduce complex behaviours into a generalised assessment tool.

The second category (“Yes, but...”), were responses indicating a recommended set of measures would be useful, but conditions were given or support for the idea was otherwise qualified. Some warned that the use of measures should not be at the expense of individualised interventions and recommendations for measures should be presented as flexible guidelines rather than mandatory rules. The sentiment expressed was to preserve practitioner autonomy and decision-making about what is appropriate for their practice and their clients, for example,

Yes – or as long as they are...guidelines and not become black/white rules to accept or decline applications [for funding].

Yes, in that it would nationalise the rating scales, measurement of care over the country, but need to maintain individual care plans etc too – a guideline [might] be more useful to give that leeway.

Although supportive, others in this category had reservations about the nature of the measures that might be recommended. Advice was given about the content, the range of clients to consider, and feasibility for practitioners.

Yes – but there is a range of age/ability etc so should take this into account. Allow for client/family...input as major part. Also should be generic for all agencies.

Yes, I think so dependent on being “user friendly”, evidence-based.

There was also appreciation of the need for measures to be useful for purposes other than informing frontline work with the client.

...information from measures should be able to be collated to address service development needs.

I think outcome measures...should be undertaken to assess if the HCN process is at all helpful in relation to the large amounts of time and money involved...

Responses in the “Yes” category were unreservedly supportive of a recommended set of outcome measures, and many reasons for this were supplied. There was recognition of the value of having a routine method to monitor progress and outcome, to standardise practice across HCN, and to save interagency teams time and effort by not having to “start off from square one each time”.

A desire to approach interventions with HCN clients from an informed perspective was also clearly present and is consistent with overcoming the barrier reported previously of a lack of awareness/availability of suitable measures.

Any information is always helpful. I want to be informed.

Yes, even if the set has drawbacks, better to start using them and getting some objective data collected than ignoring this area.

Ideal outcome measurement practice

Preparation required

Participants highlighted practical issues, contextual factors, process, assessment and measuring-specific issues in relation to groundwork that should happen before using measures with HCN clients.

Practical issues described by participants included having adequate funding and training to use measures. Identifying the appropriate time and setting for administering measures was also deemed important.

A number of issues related to the client and the wider context were suggested. Many mentioned that ensuring the client was in a stable living situation was a crucial step. In addition, agreement and buy-in from the child/family regarding the processes to be adopted was important. Also vital were inter-agency/discipline coordination and collaboration and the establishment of specialist support. Many participants

commented on the importance of adequate relationships between professionals. For example,

Agreement on “diagnosis” and on potential for change, together with commitment to shared working between agencies as to [the treatment plan].

Acceptance of expertise of each agency involved in its own field but good understanding of what each professional does – and of organisational boundaries.

The majority of responses included reference to process issues, especially the importance of having a relationship with the client prior to using measures. Comments highlighted necessary first steps as being for practitioners to establish trust by being supportive and nonjudgmental.

Building rapport with the client...once children have been in the system for some time the issue of “trust” becomes paramount – you cannot “do things” to children – they need some ownership and support.

Positive engagement with the client/family...assist[s] in the understanding and accurate representation of outcomes. I have evidenced misrepresentation of events due to lack of engagement and historical difficulties between professional and family.

One participant remarked that for some clients extra assistance should be available, for example someone to read/write for them. The ethical difficulties of working with complex young people were also raised, resolving informed consent and privacy issues in particular.

Like process issues, the central role of a thorough assessment was often mentioned. The emphasis was on the breadth and depth of the assessment at the very beginning. Additionally, a baseline period was needed, and clear goals set by the team in order for outcome measurement to be meaningful later.

Measuring-specific issues were mentioned several times. These were teams actively deciding that measures were needed, which ones to use, and ensuring they were linked to the goals that were set. Many of these comments seemed to echo those about building and maintaining effective inter-disciplinary collaboration, for example,

Discussion and agreement of agencies to use of measures, common understanding of usage, terminology etc and involvement of all agencies in use.

Domains where measurement would be useful

Participants were asked a “Yes-No” question about whether outcome measures would be useful for various domains of functioning (see Table 4). With the exception of “Physical health”, over half the participants indicated measures would be of use in each domain. More than three-quarters thought measures would be useful to monitor educational and social functioning, safety, mental health, attachment/belonging, and family relationships.

Table 4 *Proportion of Participants Indicating Outcome Measures Would be Useful in Each Domain of Functioning*

<i>Domain</i>	<i>“YES” responses n = 59</i>
Educational	86%
Social interaction	83%
Safety, security and/or risk	79%
Attachment/Belonging	76%
Mental health	76%
Family/Whānau relationships	75%
Development	68%
Independence	64%
Communication/Speech	59%
General wellbeing	56%
Cultural identity	53%
Physical health	36%

Ideal approach

In an open-ended question asking what they considered the ideal measurement tool(s), few participants actually named outcome measures. Most described an ideal approach to measurement or ideal conditions to support assessment or progress monitoring. One theme for ideal conditions was collaboration and support from all involved. For this to occur successfully, comprehensive multidisciplinary assessment, followed by

ongoing update meetings, progress reviews and monitoring of interventions were needed.

Participants felt that, ideally, measures needed to be used in conjunction with other approaches, across domains, environments, and perspectives. The emphasis was on an approach that was appropriate for individual clients. However, a number of participants thought the necessity of an individualised approach precluded routine outcome measurement. Some felt there was no ideal measure and a combination approach, integrating information from multiple sources, was needed. Outcome measures were seen as a source of only one type of information, and limited information at that, for example,

Questionnaires and formal measures can be useful but very limiting only focuses on one dimension. Also people tend to be subjective in their answers to questionnaires and have to limit their answers to the scales provided.

Specific measures that were named included the HoNOS-CA, *Outcome Rating Scale* (ORS), CORS, *Session Rating Scale* (SRS), Child SRS, *Child Depression Inventory* (CDI), *Risk Estimation System*, *Overt Aggression Scale*, and the *Head Injury Behaviour Rating Scale*. However, these suggestions were often accompanied by an acknowledgment of the narrow applicability of the content or target group, or other limitations, such as sensitivity to change.

I like HoNOSCA as this is quick to administer and process, however, those...who access HCN funding need...tools that will pick up small positive changes over time.

Overall, the nature of responses to this question showed a lack of understanding about what an outcome measure actually was. Many named assessment or diagnostic measures, which are not suited to routine outcome measurement. The most frequently mentioned were the CBCL, *Teacher Report Form* (TRF), WISC-IV, *Family Adaptability and Cohesion Evaluation Scales* (FACES), and *Canterbury Social Development Scale*. At times, a preference for the use of assessment measures over outcome measures was expressly indicated. In addition, others named techniques or non-standardised approaches that, again, were more suited to assessment, diagnosis, or as a therapeutic tool (e.g. Q-sorting cards, Family and Social Circle diagrams).

Some participants suggested general characteristics for measures they would like to see used. They asked that measures be simple, have a broad coverage of domains

including satisfaction with services, incorporate the viewpoints of all involved, be useful for monitoring progress, and be suitable for pre-post comparisons.

Recommended outcome measures

Participants were asked to recommend outcome measures they knew of or had used. They estimated the administration time, gave a reason for its use, and described the advantages/strengths and limitations/weaknesses. Eight outcome measures, 18 assessment/diagnostic measures, and 10 other (or informal) tools were named.

Once more, misunderstanding of the term “outcome measure” was evident in the responses. For example, the WISC-IV, ABAS, *Probe*, and *Conners'*, were all recommended despite being unsuitable for routine outcome monitoring, for example having administration time estimates of up to six hours. Many suggested informal or unstandardised methods, or those developed for individual use, such as Individual Education Plans, Visual Analogue Scales, daily points charts, professional meetings, and goal achievement. The eight outcome measures that were recommended are presented in Table 5, along with a summary of participants' comments about the strengths and weaknesses of each.

Table 5 *Participants' Comments about Outcome Measures They Recommended*

<i>Outcome Measure</i>	<i>Estimated Admin. Time</i>	<i>Reason to Use</i>	<i>Strengths</i>	<i>Weaknesses</i>
BDI Beck Depression Inventory	30 min	<ul style="list-style-type: none"> To understand person's internal state or depression symptoms 	<ul style="list-style-type: none"> Easy to read Quick reference point to emotional state Quick reference for safety/risk 	<ul style="list-style-type: none"> One-to-one administration Psychologist interpretation required Takes time to process
BYIs Beck Youth Inventories	20 min	<ul style="list-style-type: none"> When behavioural difficulties are present or when indications of mood problems 	<ul style="list-style-type: none"> Easy to administer Good to integrate with other data 	<ul style="list-style-type: none"> Self-report biases
Cage & Kessler Screening Tool	10-60 min	<ul style="list-style-type: none"> To indicate crisis/pre-crisis To measure level of risk To provide information about specific areas/needs 	<ul style="list-style-type: none"> Provides a snapshot picture Social workers can administer 	<ul style="list-style-type: none"> Covers limited areas Mood section is not easy to follow Not good for routine use Other outcome measures would be more suitable Taken alone, results can be misleading
CDI Child Depression Inventory	30 min	<ul style="list-style-type: none"> Can show a reduction in symptoms over time Indicates depression Indicates risk of self-harm 	<ul style="list-style-type: none"> Is self-report Measures treatment effectiveness 	<ul style="list-style-type: none"> American language Doesn't apply to very young children

CORS Child Outcome Rating Scale	< 1 min	<ul style="list-style-type: none"> • Highly feasible 	<ul style="list-style-type: none"> • Can have parents etc complete too • Feasible • Gives client rating, "which research says is a better indicator than clinician rating" • Readable • Valid 	<ul style="list-style-type: none"> • Not as reliable as other outcome measures, such as OQ-45
HoNOS-CA Health of the Nation Outcome Scale for Children & Adolescents	5-25 min	<ul style="list-style-type: none"> • Agency requirement; mandatory • Can use at beginning, middle, and end • Standard measure of change in key mental health factors • To assess progress • To collect statistics for Ministry of Health 	<ul style="list-style-type: none"> • Comprehensive; covers multiple domains • Fast to rate • Has computer scoring and graphs/reporting • Provides service statistics • Well researched 	<ul style="list-style-type: none"> • Annoys clinicians • Does not cover cultural identity • Is a weak and ineffective tool • Is generalised • Is not enough on its own • Is technical • Is useless • Lacks research • Not "valid" for use with children • Poor reliability • Psychometrics are "a bit rough and nasty" • Relies on clinician impression • Requires computer access • Requires ongoing training to ensure accurate completion • Subjective as based on clinician ratings • Training and experience required
KORS Kaupapa Outcome Rating Scale	< 1 min	<ul style="list-style-type: none"> • Can use every session • It's a "speedometer" rather than a "speed camera" 	<ul style="list-style-type: none"> • Feasible • Reliable • Valid 	<ul style="list-style-type: none"> • Not suitable for some non-Māori
SDQ Strengths & Difficulties Questionnaire	5-10 min	<ul style="list-style-type: none"> • Can be used at entry, review, and discharge 	<ul style="list-style-type: none"> • Brief • Broad measure • Easy to complete • Includes client's perspective • Includes parent and teacher perspective • Reasonably user-friendly • Well researched 	<ul style="list-style-type: none"> • Lacks detail

The HoNOS-CA was the most frequently mentioned followed by the *Cage & Kessler Screening Tool*, which are mandatory measures in DHB services and CYF, respectively. Interestingly, the impression was that neither were preferred outcome measures as participants highlighted more weaknesses than strengths and some actually stated the measure should not be used. The remaining six outcome measures were mentioned by less than three participants each.

Additional issues

At the end of the survey was the opportunity to raise additional issues to be kept in mind when outcome measures were being reviewed and selected in this project. Most participants offered suggestions, which included advice about the selection process, the characteristics of an appropriate measure, and the impact of using measures on the HCN team and their intervention. In addition, some participants provided cautions or elaborated on why they thought using measures would be inappropriate for their practice.

Cautions and reservations

The issue raised about measures not providing the full picture, appeared to reflect practitioners' concerns that this would disadvantage or misrepresent their clients or practice. The following quote was representative of the views of several participants,

More open-ended questions rather than scales with numbers. They are too simplistic and the whole picture needs to be looked at. Another risk of only using the scales is the person maybe having a bad day. Could test differently in the same week.

A large number of responses seemed based on an assumption that measures were complicated, required technical knowledge or training, or are only for psychologists. This led to calls for adequate training and education, an acknowledgement of the current shortage of skilled staff and the increased workload that would result if measures were introduced.

...the team often struggle with the concepts of outcome measurement. LSCs need to have a clear understanding of outcome measurement to ensure appropriate data is collected from specialists...

Currently, controversial because it may imply that psychologists will be expected to carry out these assessments. Think it is the best way to go, but will add to already heavy case loads

Some expressed the view that recommending specific measures was an affront to practitioners' expertise, autonomy, and professional judgment. Outcome measures were seen as unnecessarily restricting practice.

Does it improve my clinical skills? Will it force me into treatment regimes that neither suit me nor the client (there are over 400 schools of therapy – be careful you don't lose some).

Allow teams to do their jobs without measuring and counting everything.

Moreover, some felt measuring outcome would impede progress with a client or in some way hinder the intervention. Outcome measurement was seen as an intensive, intrusive process and unnecessarily repetitive.

With a focus on measures, they can become narrow...the danger is these children can have multiple assessments by many different clinicians which in itself becomes abusive, and doesn't contribute to their welfare.

Finally, there were cautions given about the potential for outcome data to be used irresponsibly to make funding decisions. Participants were worried services or clients would be reduced to simple budget items and the individuals involved would become invisible.

Advice about measure selection

Suggestions for ideal characteristics to seek in outcome measures were the largest type of response to the additional issues question. In general, participants indicated measures should be appropriate, relevant, and easy to use. A strong theme about the need for outcome measures to be user-friendly and very quick to administer, score and interpret was apparent, as highlighted in this comment,

Mental Health already involves vast amounts of paper work. Any measure you were considering using would have to be brief and easy or clinicians wouldn't do it.

A second theme was to ensure robust psychometric properties. Participants thought measures should be reliable, valid, and relevant for the New Zealand context, preferably being standardised for our population and having local normative data. In particular, cultural, religious, and developmental issues were identified as important factors in determining the relevance of the measure for New Zealand. One participant suggested a Te Reo Māori version of the selected measure be made available.

Although participants gave advice about ensuring measures were suitable for their clients, they also thought about outcome measure use at a service level. They voiced concern about the appropriate use of resources and showed preference for a measure that was inexpensive or free to use. Given that multidisciplinary teams are the norm with HCN clients, participants wanted measures that could be used by any practitioner and assist communication and planning between disciplines.

Having the opportunity for multiple informants (including the young person) was suggested, and even assumed, in many responses. Eliciting the perspective of important adults in the child's life was seen as crucial. However, a low reading age for measures was recommended to suit caregivers who may have limited education. In terms of measure content, participants emphasised social aspects of the child's functioning, such as interpersonal behaviour, school and home functioning, the whānau environment and cultural history. The diversity in presenting issues of HCN clients was highlighted and linked to the breadth of coverage any single measure would require. Participants also reiterated that measures not be the sole source of information about client progress.

Advice about implementation

Participants offered suggestions about the process of selecting and implementing routine outcome measures. This included a plea to obtain feedback from the practitioners who would be using the measures, offering training to ensure consistency across teams and services, and to build a flexible system. A few described the idea of a "catalogue" where the recommended measures in a range of domains are presented, and can be selected, as appropriate, for individual clients by their HCN team.

Summary

The *Survey on Outcome Measurement* was sent to practitioners who had worked on an HCN interagency team and 72 completed surveys were returned. The purpose of the survey was to elicit the opinions of frontline practitioners about what is currently done to measure outcome and what is considered ideal, to find out which measures they knew of, as well as provide an opportunity for additional issues to be raised regarding the selection and implementation of routine outcome measurement.

Characteristics of participants and their clients

Despite being highly qualified and experienced in their roles, most participants had little experience working as part of an HCN team. This is not surprising given the individualised establishment of HCN teams to suit each client. In terms of introducing outcome measurement, this suggests regular repeated training sessions would be required to ensure practitioners new to the HCN process have the necessary skills. A

consequence of this high frequency training is that as few resources (time, people, cost, and equipment/materials) as possible would need to be used to make it feasible. In addition, as predicted, consideration of the wide range of professional backgrounds and work settings of the practitioners will be important in the measure selection process.

The characteristics of HCN clients the participants had worked with were diverse. Presenting issues related to aggression, learning, care and protection, attachment, intellect/development, and mental health were each rated as significant in over half the responses. This highlights the need for an outcome measure that is broadly applicable, but to comprehensively cover all domains would require quite a lengthy measure. Therefore, to be feasible, a measure would include key aspects of as many of the significant domains as possible. This would enable the measurement of outcome within a routine protocol for the wide variety of clients described.

Participants' current practice

Overall, participants reported little experience using measures with HCN clients and the term "outcome measure" was not widely understood. Despite an explanation/definition being included in the survey, many equated the term with *any* method of gathering information about a client, formal or informal. Those who were familiar with the term saw the role of measures in monitoring progress. They commonly reported working in an organisation with a routine outcome measurement protocol in place, or had "experts" on the team who were responsible for their use. The idea of outcome measurement being for "experts only" was repeated throughout, which was consistent with the general impression conveyed that measures were complex and difficult to use.

Participants rated the importance of a range of indicators for decision-making about client progress and readiness to transition back to mainstream services. The most important indicators for both purposes were almost identical, and measures (self-report and other types) were among the least important. Indicators that relied on practitioner observation and judgment were rated as more important than other – potentially more objective – indicators, which was consistent with the findings of previous studies (Garland, et al., 2003; Hodgetts, et al., 2005). This was interesting, as participants commented on the subjective and vague nature of outcome measures, suggesting they prefer to avoid such bias in their evaluations. The implication is that professional impression was appraised as being a more objective source of information about client

outcome. This is consistent with literature on clinical judgment (Hatfield, McCullough, Frantz, & Krieger, 2009) and the power of cognitive biases and heuristics in social perception (Myers, 2009). This points to the vital role of education and training in the implementation of routine outcome measurement to ensure practitioners are using indicators that are supported by literature and best practice, such as psychometrically robust outcome measures.

Given that participants felt measures had little importance for their practice, a reasonable prediction would be for barriers to be more prominent than reasons to use measures. In contrast, the results were supportive of the use of outcome measures. Participants rated the reasons to use measures highly, including as a means to track change. However, reasons such as providing a check on decision-making and promoting discussion with the client were the least important. This appears to fit with their emphasis on valuing professional judgment. This was apparent in the ratings of the biggest barriers too. For example, the view that measures can be misleading out of context supports the belief that measures are subjective or preclude the use of clinical judgment. Another barrier rated as important was that measures are not user-friendly, which was also strongly expressed in narrative answers to later questions. Generally, these ratings suggest participants see the value in measures, although perhaps not as an integrated part of the therapeutic process. The findings of this study echo those of Garland et al. (2003) and provide additional support for their recommendations for services wishing to introduce outcome measures. In particular, further training could improve clinicians' understanding of the use of measures by showing how measures can facilitate rapport building, progress, and goal setting, in addition to serving a monitoring function. Minimising other barriers to their use could also be done through training, (for example, how to interpret and report results appropriately using contextual information) and during the measure selection process by ensuring measures are user-friendly, easily available, and have specific characteristics that make them suitable for use here.

An encouraging result was that most participants thought a recommended set of outcome measures would be useful for their practice in a range of domains. However, there was concern about standardised measures not capturing the important aspects of individual cases. Therefore, practitioners wished for measures to have sufficient breadth, include client and family input, cover symptom and social functioning changes, and help them work from an informed perspective. Participants could see the multiple

levels at which outcome data could be used, although some gave admonitions to ensure it did not become the sole basis for decision-making by management/funders.

Participants' ideal approach

Participants felt that prerequisites for using measures included adequate training and time for administration, scoring and interpretation. In addition, rapport with the client and a stable living situation, good interagency relationships, and resolving ethical issues were important. Many explained that before using measures a comprehensive assessment with baseline data should be completed, so goals can be set and the measure used as the intervention progresses. This showed an understanding of the processes of outcome measurement and the importance of integrating measures into routine practice.

The majority thought measures could have a useful role in all but one domain of functioning. Approximately two thirds of participants did not feel monitoring physical health outcomes with measures would be useful. Despite this endorsement of their usefulness, measures were rarely named as part of participants' ideal approach to outcome measurement. Instead, the importance of process issues, such as interagency collaboration and involvement of the client, were reiterated. When measures were suggested, it was to be part of a larger monitoring process that was individually tailored to the needs and circumstances of each client. Ideally, participants wanted measures that had a broad focus, obtained multiple perspectives, and assessed status in a number of domains. Also, even though psychometric quality had previously been rated as important, many informal/idiosyncratic techniques and tools were suggested as ideal. Use of these poses difficulties for wide scale implementation and the gathering of valid and reliable outcome data.

Participants' recommendations for measures

Most measures recommended by participants were assessment/diagnostic measures or informal techniques that were not suitable for large-scale routine use. In all, eight outcome measures were named and, overall, participants were positive about the BYIs, the CDI, the CORS, the KORS, and the SDQ. The HoNOS-CA was frequently mentioned, but was generally reviewed unfavourably. The strengths of the measures noted by participants indicated they saw the following characteristics as attractive,

- Short and quick to use
- User-friendly

- Multiple informants
- Good psychometric properties and research backing
- Wide coverage of domains
- Gives useful information when scored

In contrast, the comments about the weaknesses of the measures indicated participants were averse to measures that,

- Require lots of training
- Are too technical or hard to follow
- Have only clinician rating
- Lack research or have poor psychometric properties
- Do not provide useful information for clinicians
- Are resource intensive (time, people, materials)
- Are unable to be used with a wide range of clients
- Require other information to be collected in order to interpret results

This is consistent with previous literature that practitioners prefer measures that are efficient and have practical benefits to their work with clients (Eggelston & Watkins, 2008; Garland, et al., 2003). It is also in line with the requirements of a good routine outcome measure summarised previously; that is sound psychometrics, usefulness, and feasibility.

Additional issues

Participants encouraged recognition of the skill shortage and increased workload to implement a measurement protocol, however for some this was based on the perception that measures were difficult to use and required specialist input. On the other hand, some expressed resentment at the implication that practitioners needed guidance in this area and felt outcome measurement could place limitations on their current good practice or have detrimental effects on interventions. These views highlight the need for adequate education about measures in order to successfully introduce routine measurement. Participants also offered advice about selecting measures, which reinforced earlier comments about ease of use, accessibility, breadth of coverage and integration with other sources of information. Finally, gathering feedback on proposed measures from frontline practitioners and consideration of providing a “catalogue” of recommended measures were mentioned.

Conclusions

A diverse range of disciplines, work settings, and client groups were represented in the survey sample. However, it should be noted that it was unknown how representative this sample was of HCN practitioners in general. Participants appeared to appreciate that most areas of client need required outcome monitoring and that measures were a useful tool to do this. However, a clear scepticism about measures was evident. Concerns were raised about the narrowness and complexity of measures, the difficulty of a single measure not being appropriate for all clients, and the information provided not being useful for practice. The strong message from participants was outcome measures need to be appropriate to individual clients for improving interventions, and without overly compromising the autonomy of the practitioner or the resources of the interagency team.

The size of the sample is the main limitation to these results. No hypothesis testing could be done to compare the responses of different subgroups, for example whether psychologists' (who have the most training in psychometric tests) views differed. A survey of 1442 multidisciplinary child mental health professionals in the USA found that the most consistent independent predictor of attitude towards measurement was the clinician's professional discipline, with psychologists being the most positive (Jensen-Doss & Hawley, 2010). Due to the small numbers of each discipline participating in the HCN survey, analysis by profession could not be conducted. With a bigger sample, this could have yielded useful information for planning education and training initiatives about outcome measurement. Given the low response rate, the views expressed by this sample may not reflect the views of all HCN practitioners or those working in other multidisciplinary settings with young people. However, voluntary surveys such as this are likely to elicit participation from those with strong/extreme opinions. This has benefits for this project, as those with the most extreme views about outcome measures will probably be those who will provide the strongest support or opposition to the implementation of a routine measure. In this way, gathering those opinions at an early stage in the project means their concerns and needs can be considered in the measure selection process. In addition, the views are consistent with the results of other surveys of practitioners (e.g. Jensen-Doss & Hawley, 2010; Johnston & Gowers, 2005), including Merry et al. (2004) and Hodgetts et al. (2005) here in New Zealand, which lends support to the validity of the findings.

Another limitation to these results relates to the misunderstanding of the term “outcome measure”. Although the term was explained in the survey, perhaps this needed to be clearer. Understanding the term was fundamental to understanding the survey purpose and questions, and different responses might have been obtained if the definition had been clearer. Questions to assess understanding about outcome measurement could also have been included to determine the influence this had on responses. Nevertheless, that practitioners find it hard to distinguish assessment/diagnostic measures from outcome measures is an interesting finding in itself. Hodgetts and colleagues (2005) also noted this misunderstanding in their survey of New Zealand practitioners. Perhaps the lack of knowledge about outcomes and measurement principles in these surveys reflects a wider issue for New Zealand clinicians. Speculative reasons for this may be a gap in training programs, a lack of an “outcomes culture”, isolation from the international outcomes community, or a societal/cultural stance that differs from other Western nations.

In sum, despite a small sample size and some confusion about the terminology used in the survey, a large amount of information about practitioners’ views on outcome measurement was obtained. Two overarching challenges were set regarding the use of routine outcome measurement in an HCN context. First, measures should incorporate a number of specific characteristics valued by practitioners, and second, attention should be paid to appropriate implementation of any measurement protocol.

CHAPTER 5 – STUDY TWO: CURRENT HCN MEASUREMENT PRACTICE

When introducing routine measurement in an organisation, it is imperative to obtain practitioner perspectives from the outset (Bickman, et al., 1998). In fact, this was requested by participants in the survey of Study One. The survey gave some initial information about current practices and views of outcome measurement, and more in-depth discussion of the issues was sought in Study Two. The results of these focus group discussions are the subject of this chapter.

With focus groups being an increasingly common technique in the social sciences (Denzin & Lincoln, 2005), it was a useful methodology to obtain the views of HCN practitioners. This involves discussion of a specific topic by a group of people and can have variable degrees of structure. The approach allowed access to participants' perspectives in a manner flexible enough to explore diverse viewpoints (Kamberelis & Dimitriadis, 2005; Krueger & Casey, 2000). Thus, a series of semi-structured, facilitated focus groups were conducted. The aims were twofold and so provided a natural division of each focus group discussion into two parts. This chapter presents Part I of the discussion⁴, where the aim was to elicit practitioners' views and experiences with goal setting, monitoring change, and evaluating intervention

⁴ In Part II, feedback was sought about specific short-listed measures for Study Four. Although only one set of focus groups was conducted, the two parts are presented here separately over two chapters/studies to maintain flow for the reader.

outcomes. First, the perspectives of the HCN Advisors, who have an overview of teams around the country, are presented. Then, the results of the focus groups with practitioners follow.

Perspectives of HCN Advisors

A key to putting the HCN Strategy into practice is the role of HCN Advisor. The advisors are the bulk of HCN staff, which also includes the Manager, Policy Analyst, Professional Practice Leader, and administrative staff. HCN Advisors have regional responsibilities and their primary tasks are to (1) guide and assist local sector managers and practitioners to apply for HCN funding, and then (2) assist teams to implement their intervention plan and report to HCN when necessary.

Although direct contact with HCN young people is not part of their work, HCN Advisors sit at a crucial interface between policy and practice. As a group, they review all teams' current intervention plans and progress reports at quarterly intervals. Therefore, HCN Advisors were in a position to comment on the current state of outcome measurement from a national, general, and managerial level. A single focus group with HCN Advisors was arranged with the aim of obtaining information about what measurement trends and issues they saw from their wider perspective.

Method

Participants

All HCN Advisors and the Professional Practice Leader were invited and, subsequently, consented to participate⁵: two men and four women. Participants had previously trained and worked in a variety of professional capacities, including clinical psychology, educational psychology, and social work.

Discussion questions

The following questions were developed to guide the discussion of the HCN Advisor focus group. The main aims were to gather information about what measurement

⁵ Because of staff movement and restructuring of roles, the composition of the group has changed since this study was completed. Therefore, HCN stated that the views expressed in this focus group might not be those held by the organisation or staff since then.

practices were seen currently and what the HCN Advisors would like to see happen in the future. With some exceptions, the questions mirrored those developed for the practitioner focus groups described later.

1. What key signs/indicators do you look for to determine whether a client is benefiting from an intervention?
2. How is the decision made to end an intervention? Or, how should it be made?
3. In your opinion, what is the best way to assess how effective an intervention has been?
4. Research has indicated that measurement of consumer satisfaction is quite different to measuring behaviour change or goal achievement. What do you think about monitoring all of these?

The questions were considered a guide for the facilitators only. The timing, sequencing, and wording were expected to be changed in response to group dynamics and the content of the unfolding discussion.

Procedure

Information sheets were sent to all potential participants in advance, and a consent form was signed by each participant before the discussion at the HCN Unit in Wellington began (see Appendix B). The focus group was facilitated by me and two other members of the research team, lasted approximately 1½ hours, was audio-recorded and transcribed later for analysis. The analysis involved the systematic identification of themes in the discussion. The reporting of the results here is drawn from the independent analysis of the transcripts by the members of the HCN Outcome Measurement Project team.

Results

As HCN are a small staff who work together closely, they knew each other well and appeared comfortable expressing their opinions. Although all participants were actively encouraged to convey their opinions, the majority of the discussion was driven by participants with a psychology background. Perhaps this was due to the strong emphasis on formal measurement in psychology training. In any case, there was general agreement regarding both what they wanted to see and what they were currently seeing in terms of outcome measurement by HCN teams. Themes in the discussion were identified upon inspection of the transcript and these are described

below. Anonymised direct quotes are utilised to illustrate the views shared by participants (Boyatzis, 1998; Krueger & Casey, 2000).

a. Current outcome measurement is problematic

During the initial discussion, a number of participants were despondent about the lack of systematic or meaningful outcome measurement being undertaken by teams. This information was considered a fundamental component of ascertaining progress and therefore guiding the ongoing management of financial resources, particularly for HCN Advisors, at quarterly reviews. There was a degree of frustration about teams not providing the outcome information that was requested or useful for the Advisors.

There was a small amount of objective data being reported by teams, but not generally the type or in the manner that participants felt were quality indicators of progress. Generally, these were seen as being too subjective, poorly matched to the intervention goals, or lacking reliability. One example,

...there have been an effort by a team to make up some scales where people are sort of doing a rating, which is light, but it allows you to compare today versus tomorrow versus six months...that could be helpful.

Participants were concerned the balance between qualitative and quantitative data provided by teams was skewed heavily in favour of narrative, anecdotal information. Despite concerns about teams' reliance on qualitative information, participants acknowledged it could potentially be useful and did "have a place" in the review process. For example,

As a [HCN] Advisor, the key signs that I look for are really an increase in their current well-being. I got a plan review yesterday and in it there was a lot of consumer feedback, feedback from the family as well as the young person. And there were photographs, so the panel could actually see an increase. This kid is really happy; there's photographs of him kayaking.

There was general acceptance of this notion and the inclusion of the young person's perspective was considered a significant omission much of the time. The form teams use to report progress included space for the young person's feedback and this section was not being completed on a routine basis. Participants thought the perspective of the client and family was generally not being obtained adequately, although it was unclear why this was.

Another feature of current measurement practice was the loss of the “big picture” for each young person. Participants explained this happened by breaking down overall well-being in a reductionist manner, as required by the structure of the reporting form. In addition, having no general measure of overall well-being meant good progress was often not adequately captured in the review process.

Related to this, participants mentioned that often the “issue” was not always measurable within the young person, as often it was part of their wider context. There were doubts raised about the usefulness of compartmentalising well-being for Māori and the importance of taking a holistic view.

Taking in, as much as possible, just overall, all their needs culturally because they fit into every domain...it's not just in Cultural Identity; it's in Family, there's a cultural component to every part...

Another way in which the big picture was lost was reporting outcome in relation to each intervention strategy. This too was connected to the way teams were asked to report back.

I think in terms of the current process, it feels like every little intervention and the steps that are broken down, there's a sense that there's an expectation to measure on every part of that and sometimes I feel that the process at the moment is too micro in its analysis...We need to take maybe a little more of a global look at things, and step back a bit?

b. Barriers to outcome measurement

The reporting form asks teams to comment on outcome in nine domains of well-being. Participants noted there was a high degree of discontent among teams regarding the form. The format was presented as a systemic barrier to quality outcome data being supplied for the quarterly progress reviews. Participants described the form as not being user-friendly so teams were more likely to make errors or miss out important information.

In addition, they mentioned that occasionally quantitative data was being collected by practitioners, but not recorded on the form. Participants said they knew formal measures, such as the SDQ or HoNOS-CA, were used with some young people, but were seen by practitioners as part of their role in their usual agency (e.g., CAMHS), not as part of their work for HCN.

Sometimes there is actually quite a lot of what I would consider to be useful outcome data, but it never actually sees the light of day in the context of [HCN] teams...we all know that people are not collecting the data, but sometimes they are collecting it, but not reporting on it.

A second systemic barrier to routine outcome measurement appeared to be a lack of culture of collecting data in a structured way for HCN. Participants thought this could be partly related to the professional disciplines present in many teams, as the use of formal measures was not included in the training of many practitioners. This was seen to lead to a particular style of thinking, which did not lend itself to outcome measurement.

...defining the outcomes, a lot of people aren't familiar with that sort of work. So so often if we're looking for something specific and measurable, for some of them it is quite a challenge to actually think in those terms...

In addition to systemic issues that acted to inhibit the use of quality outcome data, participants also identified other barriers. The competence of practitioners to administer psychometric measures, as well as access to suitable measures were raised. Participants noted that when formal outcome data had been reported it was often from education and psychology practitioners. They stressed that while they wanted to recommend measures for teams, they also wanted to ensure practitioners were not practicing outside the scope of their training. They indicated agencies were often under-resourced and did not have ready access to measures.

Participants also reported facing a degree of resistance, both direct and indirect, to quantitative outcome measurement and this existed at several levels. They described teams/practitioners as having little understanding or desire to implement outcome measurement because this was not consistent with their professional paradigm. They also indicated it was often even more challenging to measure outcomes reliably with the family unit because it was not consistent with the way many families operate. To illustrate,

A: I mean, sometimes you've just got to keep data, and people say, "But that's not what family and whānau do."

B: And it's really hard to step in as a practising psychologist and say–

A: –Oh, and you still get mum and dad come in and say, "Oh, well, we had three wobblies this week". How big was the wobbly? You know? Was it a really shaky wobbly or was it threw a wobbly and it landed on the

ceiling, you know? So that's a very, very, very basic sort of issue that we can't necessarily rely on people to be good sources of observation, the rigour that goes into observation, systematic, you know we need to keep some records of these things.

X: Yeah, and like, trying to explain it as well, like, "What do you mean this is a way of measuring outcome, or as a way of tracking a young person's well-being, or seeing how they are getting on with life? Really?" You know, and having to justify that stuff, it's like, "Nah, I'm not going to do it", eh.

c. Perceptions of the HCN teams' views

There was a strong opinion among the participants that teams view outcome measurement as simply a bureaucratic exercise that is not part of their work with the young person. These comments were endorsed by the group,

...well, we've got to run across our columns [in the reporting form] and you've got an intervention and a resource and an expected outcome and you've got to measure it and it's almost like getting a measure for the sake of getting a measure, but it doesn't actually add a lot...

There should be a feedback loop...that this isn't something that you do just with HCN.

Participants also identified the lack of link between reviews of progress and the original goals, which reflects a further feedback loop but internal rather than external to the team. In addition, they were concerned about teams' ability to write goals that effectively guide interventions and facilitate reflection on change. In this context, teams were presumed to have little knowledge of the concepts of measuring outcome. Participants also commented that teams were suspicious about outcome measurement. At the same time, they were concerned themselves about the implications of not being able to show HCN is a useful way to spend government money, for example,

Anyway, there'll be a consequence if we can't show that we're doing something useful with this resource, we could lose it...

d. Ideal outcome measurement

The development of a "catalogue" of measures HCN Advisors could then use for recommending to teams was suggested. This would ensure they were able to advise

practitioners about how to improve outcome measurement for their interventions. They emphasised the importance of measures being appropriate for a range of practitioners and the working environment of teams.

A: ...what I think would help my job as an advisor to the teams that I'm supporting, is to know what assessment scales, checklists, outcome measures are currently in good use, used well–

B: –good reputation–

A: –and what they are best used for, being very careful not to misuse tools...so if we could have some, like a menu of already existing scales, checklists, outcome measures, and descriptions of the population for which it is most useful...and an idea of whether you have to be a psychologist, a psychiatrist etc. So we're not saying to someone quite inappropriate, please do this...

B: And how practical are they. If you think of really busy people, is this something that's realistic?

However, participants also recognised the need for outcome data at several levels, and discussed the idea of a routine measure that was used by all teams so outcomes could be aggregated. They acknowledged that doing this in a standardised manner is complicated by the diversity of HCN young people, but indicated that gross positive progress could still be tracked, even with varied goals and interventions. The following conversation illustrates understanding of the essential idea of routine measurement by using a “good fruit” metaphor.

A: The problem is sometimes we're looking at apples and oranges when we're looking at outcomes to say what is going on for this young person versus this young person, and having some common tools and we feel that there's a need to standardise that...maybe the concept of having a standard form would even be not so helpful...

B: You could say apples and oranges, but we all know what good fruit is. So, I think that what we need, clearly, is going to be a mix of individual measures that can track and assess individual progress, and then maybe some lowest common denominator stuff, but then some way of aggregating the individual improvements...

A: So, you're looking for a good fruit measure, and the way you get good fruit, like for this young person, is you've got to teach them academic skills...and for this one, it's some different things...

Participants also talked about how building understanding and capacity, with regard to outcome measurement, in the HCN workforce was vital, and included themselves in

this. They wanted to be more informed about outcome measurement to give better advice to teams.

Participants explained that they wanted to track improvements in well-being, but this was not defined further than the HCN well-being domains, which they had already identified as problematic for this purpose. However, they noted that the family relationships domain currently suffered from having the least rigorous data collected. In particular, the importance of positive attachment for HCN clients was discussed at some length. This was seen as often underlying a range of other common presenting issues and a measure of family relationships or attachment was considered desirable.

This was linked to the value of incorporating whānau well-being in interventions, to enhance Māori young people's connection to whānau. There were also calls for outcome measures for HCN to be appropriate for our bicultural context.

...ensuring that the tools are going to be Māori-friendly, sorry that's not very good terminology, but it's really important that any measurement tools, that they are going to invite participation...it's just about how they're designed, the language that's used, the target audience, all of those things are going to be really important.

Finally, there was discussion about participants' collective view that the quality of collaboration in the teams was a key predictor of positive outcome for the young person. There was a desire for this to be monitored in some way, although HCN were moving to do that independently of this project.

Summary

A semi-structured focus group was conducted with HCN staff about how outcome is monitored by frontline teams at present and how they thought this could be improved. The discussion of these issues flowed easily, with the group appearing to agree on the issues raised by individual members.

With regard to the current practices HCN saw across teams, the consensus was that it was far from ideal. Participants were concerned about the lack of systematic data collection of any kind, but particularly of a formal, quantitative nature. Most outcome reporting was in the form of frequency of proxy indicators (e.g., school attendance), idiographic informal rating scales, and most often, narrative descriptions of progress,

and summaries of feedback from providers and caregivers. Child self-report was not included by most teams and this was raised as a major concern.

Two related systemic barriers were identified as inhibiting quality outcome data collection, the way progress and outcomes were reported, and the lack of an outcomes culture. The progress reporting form teams are required to complete for HCN was cited as not being conducive to good quality outcome measurement. Primarily this was because it encouraged “micro-level” analysis, which happened in two ways. First, teams are asked to report on outcome in separate domains of well-being and this compartmentalising was thought to lead to a loss of the big picture and be inappropriate for Māori. Second, teams report outcome for each of their intervention strategies, which again meant that overall progress was not being adequately captured. In addition, the form was thought to be user-unfriendly and responsible for the non-reporting of the small amount of quality outcome data that was being collected. The other systemic barrier was the lack of a culture of outcome data collection and reporting. HCN felt that practitioners do not see what they do in their usual agency as being transferable to their work on an HCN team.

Additional specific barriers were the competence of practitioners to use measures; access to suitable, sound, and brief measures; and resistance from practitioners and clients. Participants said they would like to recommend formal measures to teams, but were mindful of the need to do this in an informed and responsible manner. That is, to know which measures were current and appropriate for use by practitioners with a range of training, experience and resources.

The Advisors felt that teams saw outcome measurement as being a purely bureaucratic exercise. They perceived discontent, misunderstanding, and suspicion about measuring outcomes among teams. They felt teams often had difficulty setting useful goals to drive the choice of intervention and did not appear to reflect back on their original goals when reviewing progress. In this context, the Advisors acknowledged the pressure they felt to provide evidence that HCN money is being spent in an effective manner.

In terms of ideal outcome measurement, the participants suggested a catalogue of robust, appropriate measures for all domains of well-being that they could utilise when consulting with teams. They wanted to have increased knowledge themselves about the topic, so they in turn could increase the ability of teams to collect quality

information. They recognised the need for outcome data at multiple levels and the notion of a standard, global outcome measure was discussed. Although concern about losing idiosyncratic progress indicators was raised, the general sense was that this could work in practice. With regard to the nature of measures, participants identified the area of family well-being and attachment relationships as important and called for measures appropriate for New Zealand and Māori.

Perspectives of HCN practitioners

Given HCN practitioners are the frontline users of measures, the next set of focus groups were concerned with obtaining an understanding of the current state and future direction of outcome measurement from them. To ensure a wide spread across the country, five focus groups covered both the North and South Islands in both main and provincial centres. The final sites were Auckland (two groups), Palmerston North, Nelson, and Christchurch.

Method

Participants

HCN provided contact addresses for all current and past practitioners who had worked in HCN teams. Information regarding the professional disciplines or other demographic details of the potential participants was not available. An Expression of Interest Form and Information Sheet was sent to each of these individuals (see Appendix B), with extras being sent to local sector managers in the locations of the focus groups. Participants registered their interest with the research team via email, telephone, or postal reply slip. In total, 18 participants attended; fourteen women and four men. They represented a range of agencies typical of HCN teams as follows,

- Three Local Service Coordinators
- Four from the Education sector
- Six from Child, Youth, & Family
- Three from the Mental Health sector
- One from the Justice sector
- One from the Disability sector

Discussion questions

A semi-structured discussion format was used. The areas covered were monitoring change and measuring outcomes in the HCN context, participants' attitudes to outcome evaluation, and their general opinion on the topic. The questions for facilitating discussion were,

1. What are the key signs/indicators you look for to determine whether a client is benefiting from the intervention?
2. How do you decide when to end the intervention?
3. In your opinion, what is the best way to assess how effective an intervention has been?
4. What would your response be if HCN asked you to begin using formal measures to monitor a client's progress?
5. Research has indicated that measurement of consumer satisfaction is quite different to measuring behaviour change or goal achievement. What do you think about monitoring all of these?

Procedure

Each focus group was conducted with a small group of participants by two facilitators, in a professional venue, with refreshments provided. A consent form was signed by each participant before the discussion began (see Appendix B). Each focus group lasted approximately two hours, was audio-recorded, and later transcribed for analysis.

Analysis of the transcripts identified recurring themes or patterns of response (Boyatzis, 1998). The aim was to give structure and order to the practitioners' views on monitoring client change and using outcome measures. The themes that appeared particularly strong and pertinent are presented here and it should be noted that there was consistency in the themes identified across focus groups. The reporting of the results here is drawn from the independent analysis of the transcripts by the members of the HCN Outcome Measurement Project team.

Results

The dynamics that developed within the focus groups was influenced by a range of factors. First, there was a relatively small pool of practitioners within each geographical area, which meant most group members knew each other and were even working on the same HCN team. Second, it was noticeable that despite the presumption that drawing from small local pools of practitioners would ensure each group was a microcosm of local skills, the groups did not generally contain much diversity. The

result was that some groups focussed more on family and welfare issues, others more on education. It was conspicuous that health and mental health issues were infrequently mentioned in most groups.

Participants were guided by key discussion questions; however, there was a large degree of flexibility. It is important when considering the discussion themes presented below, to bear in mind that the groups appeared keen to ensure the facilitators understood that HCN clients were difficult to engage and work with, had (by definition) complicated needs, and that substantial change was unlikely to occur within the 1-2 year timeframe usual for HCN funding. Participants were worried HCN may choose to implement an outcome measurement protocol that was geared to the needs of a “higher functioning” cohort without *really* appreciating the complex needs of the HCN group, as illustrated here,

Christchurch

I also think the HCN team and the [Advisors], I think they sometimes need to be aware and meet these real people. I do know the process, but we've got another life and another role, which is far removed, quite often, from what is on [the HCN intervention] plan. I always remember the girl we had for three years...Anyway, it was just before Christmas and I thought, these people in Wellington – cos that's where they all were – making these decisions and I invited...a guy down. I said, “If you're calling in, pop in to the unit and meet this girl”. I tell you what! Things moved dramatically when they saw the person. She was very difficult, there were no two ways about that, but there was a reality in them seeing the person; they're not just, “Jack Smith”.

A number of broad topics were covered by the focus group discussions, (1) spotting change, (2) methods and tools, (3) accountability, and (4) monitoring and management. Of course, although the talk across groups is presented here as a series of broad topics and themes within these, it was not neatly packaged in this way during the discussions. Representative direct quotes are used here to illustrate key themes covered by the participants.

a. Spotting change

What leads to change

Some participants felt the focus of any attempt to track change should be based on those factors that influence the change process rather than attempt to assess each

detailed outcome. This position was present in most groups in one form or another. The general form was to place much less emphasis on the psychometric measurement of specific “outputs” and more on those aspects of individual functioning that underpin such outputs, despite these being much more difficult to measure. This view appeared to reflect the focus of current practice on intangible factors, for example “respect for others”, “positive relationship with parents”, “working hard at school”. It seemed to be exactly the difficulty that could lead to intervention plans being vague and placing greater emphasis on inputs rather than outputs, such as “providing better access to appropriate peer groups” without specifying the desired result of this, or “providing appropriate accommodation” without specifying what the new arrangements will achieve that the previous arrangement could not. It was observed clearly in one group that such an approach would easily result in the specification of strategies rather than outcomes, as pointed out by a fellow participant in the following excerpt.

Auckland 2

A: One thing we do with the family is...with the young person who is with the caregiver, we've got goals around her knowing her family tree and finding out about her family history and we can definitely measure those sort of goals.

B: But that's not...that's a strategy; it's not really a goal is it? What do you want to achieve from that?

A: We want to achieve...it's about her identity, that she has an understanding of...she really doesn't know...she knew she had a [family member] who died in [NZ town], so we want her to know about her family history and believe that she has somewhere where she's come from.

B: So then, the goal is really about finding out about her identity and the strategy is to do the family tree, but how do you measure that she's found her identity?

Despite some very interesting discussion in each of the groups on the subject of change there was no clear view expressed on how to identify when this has occurred. That is, there was no predominant theoretical model being applied to the work being undertaken for HCN, even by workers employed within the same agency or from the same professional discipline. A number of participants referred to the HCN domains of well-being as providing a framework within which they conceptualise their work. The perception was that this was of little value when incorporating formally evaluated outcomes into an intervention plan because a number of domains do not lend themselves easily to this style of measurement.

Setting goals and identifying targets

There was a theme concerned with using the goals from the initial intervention plan to provide a focus for outcome measurement. However, a problem reported by a number of participants was that their goals were often not well formulated and were not amenable to observation or evaluation, as described here,

Auckland 2

When I first joined the team, I found the goals and the outcome measures to be really airy-fairy. I came in at a review time, so I sat there through the transition when I sat there with the other psychologist, and the whole team didn't know where to start. They had this plan and they said, "Well, how's she done?" The goals were so immeasurable they didn't even know how to go through the plan...They had this conversation for about 50 minutes because the outcome measures were so vague, like, "X will be happy." Well, how do you know if she was happy?

This quote illustrates a number of additional points related to tracking change using a process of goal setting and review. First, the success of any plan, or at least the isolation of evidence to demonstrate change, depends on how well the goals are selected and drafted. In this example, the goal ("X will be happy") does not fit a specific domain of well-being, nor is it clear how it would be evaluated in an objective sense. Second, unless goals are considered and specified at the time of designing the intervention, the task is time consuming, and perhaps impossible, at the time of review. These general points were reiterated in other groups. The following quote illustrates another factor to be considered when drafting goals: adherence to the format that will be funded, which is dependent on doing what can realistically be achieved in the timeframe. This sentiment was expressed repeatedly.

Nelson

To be really frank...because funding requires outcomes to get the funding, so I have a very very pragmatic style in that when I am drafting up the intervention plans on behalf of the teams, we work to make sure that whatever goals we set and outcomes we think we're going to achieve, are achievable, so we don't go setting out something that in three months, six, twelve months time, we think we're going to fail on. It's quite a pragmatic approach to say what can we realistically expect to achieve, and sometimes it might be something as specific as that a particular specialist input be completed, like the specialist input will be done, the training will be done, a behavioural management plan will be

in place and the family will be able to describe what things they are doing in the home environment.

Throughout the conversations, there seemed to be a range of understandings about concepts as apparently straightforward as “goals”, “intervention” and “planning”. Participants (as above) cited typical goals, which were focussed more on processes or strategies rather than change or desired results for the young person’s well-being. There was a lack of common language regarding what was a goal, an outcome, input, or intervention, across participants.

Secondary indicators

A number of participants referred to secondary factors that would indicate change had occurred. These included a reduced need for additional funding, reduced supervision (e.g., teacher aide time), or reduced contact with services (e.g., police, CYF). These indicators are considered secondary because they are subject to a range of influences in addition to the young person’s well-being. For example, reduced teacher aide time could reflect changed funding priorities rather than being a direct result of behavioural or educational change. One participant noted the secondary nature of some formal outcome measures,

Auckland 1

[Psychometric rating scales]...they’re once removed. The prime indicator is the change in behaviour.

This implied a need for outcome measures to capture the change in the young person as directly as possible. That is, by documenting behavioural change, rather than the results of behaviour change, for example, service usage.

b. Methods and tools

To use or not to use standardised measures

This theme addressed the use of formal and standardised systems of outcome measurement. A strongly held view across groups was that while HCN may want to implement some norm-referenced measures as part of the overall intervention planning and review process, there were concerns about the practicality of this. For example,

Palmerston North

Facilitator: I’m wondering how, if you were asked by HCN to start using some standardised measures, how that might go down?

A: It wouldn't. I can't imagine it would go down very well. It would be more paperwork that we'd have to do to review a system, at the end of the day the children are high and complex.

The view was expressed that no single measure was available that was sensitive enough to the individual needs of young people who receive HCN funding. Also, there was a sense that such monitoring was of peripheral concern to the main task of setting clear goals. That is to say, at best, formal outcome measures were an add-on and there were reservations about the value of quantitative methods of tracking change. One participant summarised this view,

Nelson

So, I just want to put in a bit of a caution thing that sometimes, because it is quantitative, because it is in a box and we can scale it a, b, c, d, it seems more valid information and I just question that, because my experience and practice is that it hasn't always meant that much, except it is a good way of securing funding.

There was acceptance that norm-referenced measures had a place in some domains of well-being. For example, for physical and mental health or academic achievement, standardised evaluation and comparison against appropriate New Zealand norms was viewed as reasonable and appropriate. However, participants felt that generally only psychologists have access to such tests and training to use them. These could be significant issues for outcome measurement implementation and might be the reason behind a number of statements made in the focus groups. One reason stated by most participants for the limited role of standardised measures, was their lack of familiarity with them as a method of tracking change.

Formal versus informal progress review

Between formal psychometric measurement at one end and clinical judgment at the other, are a wide range of semi-formal and criterion-referenced tools and techniques for monitoring change. Such techniques may not give information about how an individual is functioning compared to their peers, but can be used to show change against a baseline or highlight interventions that are most likely to have promoted this change. A semi-formal individualised form of monitoring was mentioned by very few participants, despite the ease with which such protocols could be introduced. One example,

Auckland 2

We've measured her behaviour change through...a points system, so she is monitored basically on an hourly basis and she gets a mark out of 3 of how she's been doing...We've been able to track her changes by how many 3s she gets in one day, or how many detentions she gets, or how many times she's received a reward weekly, so it gets monitored in a quantitative way.

As mentioned above with regard to difficulties drafting goals, even when goals were developed, the methods with which they are evaluated were not. This was not simply the absence of formal monitoring tools, but also the absence of informal ones.

Obtaining “buy-in”

Many participants felt that whichever strategy was used, it needed to be broad and multi-faceted. Crucially, the strategy should also be able to secure buy-in: the cooperation of the client, family and, presumably, practitioners too. This was a strong theme across groups. From Christchurch for example,

I don't think it matters what [measure] you use. It's about getting the buy-in from the people and the reality is people want to know what they're going to get out of it, what is going to be the result. I think you could almost use any of these forms...but it's how you get the buy-in.

The next comment reflects the trepidation felt by many participants regarding routine outcome measurement. It is a plea for a slow rollout of recommendations that emerge from this project.

Nelson

I have one request from today really, and that is, whatever the outcome measures are, please, please, please do a trial period, rather than just going across the country with everybody at once and if it can be done in a graduated way, rather than just this whole thing all at once, to say “let's start building it in, in a series of steps”, but to have a trial and see how it works.

c. Accountability***Querying the present project***

While there was a degree of support for introducing a more robust framework for measuring outcome, there was also a degree of cynicism about the “real” reasons for this. This appeared to be due to concern about what would drive the selection of

measures and how information from them would be used. A number of participants certainly did not feel the motivation behind this project was to improve their work with clients. For example,

Christchurch

Trying to evaluate the human interaction stuff is really very very difficult and I suspect this process is really just to show that they're actually accounting for their money well. That's probably very cynical [Group comments] yeah; it's probably actually what the truth is.

Auckland 1

To my eyes, [using outcome measures] is them trying to control us....

Working within HCN structures

There were comments related to the commitment practitioners had to the wider community and the families of young people with whom they work. In particular, participants expressed concern about overburdening families with monitoring and measures that are not useful. However, they recognised that these same measures could provide an avenue for engaging the family in the change process. It was conspicuous that very few comments were made about using measures to engage young people, or that young people might find completing measures useful or interesting. The comments suggested participants saw outcome measurement as an addition to the intervention rather than an integral part of it. That is, the context for using measures was provided by HCN's requirements and not by the intervention plan.

d. Monitoring and management

The contribution that outcome measurement could make was emphasised by some participants, although some reported using measures independently of the HCN review process. This indicated a discrepancy between outcome measurement for management or funding purposes, and for clinical feedback or decision-making. The view was clearly expressed that HCN provide funding, and teams decide which services are required and how to implement these. As a result, outcome measures could be used by the team, but it was not seen as necessary to supply this information to HCN.

Christchurch

It's not an evaluation I see we'd be providing to anyone else. For me it would be something that would be guiding us looking at our original goals in the plan and saying how they are working, what might be the next step, because, "gee, for all our efforts, he still feels as if he's

hopeless” and we’re thinking that’s an area we need to incorporate more or look at in more detail in our plan, not something that we would send to anyone, saying “we’re doing okay, our outcomes are 90 percent da da”, so it would be like guidance on our own plan.

e. Feedback from the client/family

Groups were asked whether they routinely collected feedback from the young person and/or their family, how they valued it, and what part they thought it could play in the evaluation of outcomes. No participant described collecting this information in a formal or standardised manner. Most felt this feedback was valuable, but only when the young person was deemed to have the capacity to provide it. They noted the importance of engaging the young person in the intervention process.

Christchurch

I’m reflecting that on most of the ones I’ve been involved with, we’ve made a big thing – partly because they are older kids – of getting the young person to be part of the plan and they sign it off...I’m thinking that as we go through the process of measuring the goals that we set out in the original plan we have talked of going back to the client young person at each point in the plan and saying...like, “What do you think about how things are going for you? What is your next step? We’re going to be meeting tomorrow and there are five of these well paid people sitting around the table, but actually it is your plan so what is it that you think?”

As the quote above indicates, there was a general belief (and experience) that seeking the views of the young person can be very helpful. However, this view was not universally held, as seen in the following dialogue.

Nelson

A: Well, these are totally antisocial kids we’re talking about...so we do it with the parent. It’s not with the child anyway, it’s with a parent.

B: You try and talk with the kids about it.

A: Yeah, but your average kid, what they would put in and then these high and complex need kids: it would just be worth nothing.

Summary

Despite repeated comments to the effect that HCN clients were the most complex and challenging on their caseload (almost by definition) most participants believed that their own experience and judgment were sufficient and superior to formal measures. While the majority of participants were able to see benefits from using outcome measures,

they did not link these directly to benefits for the client. It may be that this was not stated because it could be taken for granted. The concept of accountability, that is, the need to check that the funding was being put to good use, was certainly understood. However, these two elements – enhancement of client wellbeing and the enhancement of accountability – were not inextricably linked in the discussions. Indeed, this might be behind the concern about HCN's motives in commissioning this project and the doubt about whether outcome measures would have any value to contribute.

Participants saw the core of the intervention as the goals established as part of the initial planning. In general, if outcome measurement added to this, it was seen as positive, but on the condition that it did not involve a significant amount of additional work. The belief that this would be the case was not widespread. In general, the groups did not discuss the relative merits and deficits of outcome measurement in a way that indicated a high level of knowledge about the area. This is not surprising given that this would not be part of the core training of many practitioners. In the main, the exceptions were psychologists who appeared more conversant with assessment theory and practice. Lack of training or familiarity with measures may have been an important factor guiding the responses of many participants. It was also notable that the relative lack of awareness of measures was not limited to norm-referenced or formal psychometric instruments. Relatively few participants mentioned criterion-referenced tools or idiographic strategies to evaluate change for their clients. Again, as was indicated by the results of Study One, taking into account the need for education and training regarding outcome measurement should be part of the measure selection process and a precursor to protocol implementation.

Conclusions

The views of the HCN Advisors appeared to complement those of the frontline practitioners. The Advisors noted that there was very little systematic, formal and quantitative outcome measurement being conducted as part of HCN interventions. The comments made by practitioners appeared to validate this. The explanations posed by the Advisors as to why this might be were also supported by practitioners who noted the difficulty using the HCN reporting form and the lack of outcome culture for HCN, despite some mentioning they do use measures. There was an implication that some practitioners did not see the information from measures as being of interest to HCN, rather they were seen as a clinical tool.

The HCN Advisors felt practitioners had trouble using goals to review progress, did not have the knowledge or desire to use measures, and were suspicious about requests to do this. Based on the conversation in the practitioner focus groups, the Advisors had an accurate assessment of the situation. Conversely, the practitioners were worried that routine outcome measurement was being considered by HCN with the intention to show the funding was being well spent. Based on the comments of the Advisors, the practitioners were right too, although this was only one reason noted by the Advisors for wanting measurement that is more rigorous. Both practitioners and Advisors recognised the practical barriers to outcome measurement, including the difficulty finding an appropriate standard measure, administrative burden, training, and overcoming resistance from practitioners, families, and young people.

The views expressed in these focus groups are consistent with previous similar research. In particular, the limited familiarity with outcome measurement concepts – which is related to cynicism about the contribution measures can make at multiple levels – has been found in research in New Zealand and overseas (Bickman, et al., 1998; Hodgetts, et al., 2005; Johnston & Gowers, 2005). Rao, Hendry and Watson (2010) reported clinicians in their service were anxious about the implementation of routine measurement for a number of reasons. They were worried about their practice being evaluated, their theoretical orientation being incompatible with measurement, the adequacy of the measures, and a lack of time and resources. It seems that the perspectives of HCN practitioners are typical of those in organisations without routine outcome measurement. Fortunately, this means that the information available on how to minimise resistance and overcome the barriers to implementation is applicable and could be drawn upon to avoid the difficulties other organisations have faced.

In their focus groups about selecting a routine outcome measure, Koch, Lewis, and McCall (1998) found clinicians valued (a) being able to use the measure for treatment planning, (b) coverage of functioning and symptoms, (c) good normative data, and (d) the measure being able to be used by a range of staff members. While these ideas were mentioned or implied by a small number of participants in the present study, they were not strong themes in the discussions. This may be a product of the low level of knowledge about measures, or perhaps were simply secondary to the concerns participants wished to express. In this regard, the findings of the present study were consistent with those of Koch et al. who found clinicians' main concerns were about cost and measurement being an extra burden. Johnston and Gowers (2005) also

reported the main obstacles to routine measurement identified by clinicians were resource issues, clinician resistance, scientific merit, utility, and perceived client engagement issues, which were all concerns expressed by HCN practitioners.

Given the small number of participants, the representation of the various sectors was unbalanced, and no information about the professional discipline of participants was obtained (although some disclosed this during the discussions). As a result, the themes that arose in the practitioner focus group discussions may not encompass the full range of opinions, knowledge and experiences, or be representative of HCN practitioners in general. In addition, no comparative analysis of participant contributions based on professional discipline could be made. Since psychologists (and many teachers) have training in assessment theory and psychometric measurement, it is possible this greater knowledge base might more favourably dispose them to the notion of routine outcome measurement compared to other practitioners (such as in counselling, social work, medicine, or justice). Despite these limitations, the results of these focus groups provided an insight into the views of practitioners and HCN Advisors, which are consistent with those described in previous studies.

CHAPTER 6 – STUDY THREE: THE SEARCH FOR SUITABLE MEASURES

The main purpose – and therefore the fundamental study – of the HCN Outcome Measurement Project was the search for a measure suitable for routine use with their teams. Study Three was the extensive search and review of outcome measures available for use with children and adolescents, which was conducted in mid-2006. Fortunately, this challenging task had prescribed parameters consistent with what might be expected for any multidisciplinary, broad-ranging service for young people. Some of these parameters were implied by the nature of the organisation and their clients, others were related to the domains of outcome to be measured, and the criteria set by HCN for their ideal measure.

Parameters of the search

The nature of the work of HCN

The nature of the HCN Strategy presented a number of challenges for selecting a routine outcome measure for teams to use and three stood out in particular. First, as highlighted by participants in the focus groups, the cases funded by HCN are extremely varied. To find a single outcome measure, or even a small set of measures, that could be applicable for all these cases, while also balancing breadth of coverage with

sensitivity to change, was an issue kept in mind throughout the selection and review process.

A second challenge was the age range funded by HCN, between 0 and 20 years old. Developmentally, young people within this range are highly diverse. The types of outcomes applicable to an infant are quite different to those of a teenager finishing high school. The requirement for the search to yield a measure applicable to all ages and developmental stages, and yet still be sufficiently useful to practitioners, was expected to be difficult to meet.

A third issue concerned the multidisciplinary nature of the HCN teams who would use the measure. As teams are comprised of practitioners from the health, education, and welfare sectors (and various professional disciplines within these), the models of practice they work from are highly diverse. The results of the survey and focus groups with practitioners suggested that their professional philosophies and theoretical allegiances might influence whether outcome measurement is viewed as an appropriate or valuable addition to their current practice. In addition, the range of disciplines represented varying degrees of training, experience and knowledge of psychometric measures and their use. Needing to consider the impact of this variation among practitioners was expected to pose challenges for measure selection and review. What might be suitable for one group of practitioners might not be suitable for another. The ideal measure would be able to be used comfortably by all, and its common language would enable results to be shared among sectors and disciplines.

Domains of outcome

The HCN Strategy lay down the foundation principles of this collaborative approach to meeting the needs of young people (High and Complex Needs Intersectoral Unit, n.d.). Seven domains of well-being were outlined, which HCN teams are required to consider when formulating individualised plans. This well-being framework was used as a starting point for identifying possible areas of outcome to be evaluated with a formal measurement tool. From the statements contained in the HCN Strategy, the key areas of outcome for each domain were identified and are summarised in Table 6.

Table 6 *HCN's Seven Domains of Well-being and Associated Key Areas of Outcome*

<i>Domain</i>	<i>Key Areas of Outcome Identified</i>
1. Physical Health & Mental Wellbeing	<ul style="list-style-type: none"> a. Illness/Condition (i.e., symptoms and severity) b. Disability/Functional status c. Sense of control d. Coping strategies e. Emotion regulation f. Management (i.e., identification & degree of needs being met) g. Health behaviours (e.g., compliance, help-seeking, engagement)
2. Attachment, Belonging & Family/Whānau Relationships	<ul style="list-style-type: none"> a. Attachment to family/caregiver/other adult b. Family functioning and relationships
3. Cultural Identity	<ul style="list-style-type: none"> a. Awareness b. Participation c. Ability to draw strength from a cultural/spiritual identity
4. Development & Independence	<ul style="list-style-type: none"> a. Developmentally appropriate skills b. Independence (or preparation for) c. Self-care skills (sleeping, eating, toileting, etc) d. Daily living skills (cooking, cleaning, travel, etc) e. Strengths
5. Social Interaction	<ul style="list-style-type: none"> a. Peer relationships b. Engagement in social situations c. Community participation d. Age-appropriate sexual behaviour e. Age-appropriate social behaviour f. Age-appropriate emotional behaviour
6. Strong Educational Foundations for Lifelong Learning	<ul style="list-style-type: none"> a. Knowledge and skills b. Language/Communication skills c. Qualifications d. Participation and engagement e. Motivation for lifelong learning f. Resilience g. Sense of identity and self-worth h. Attitudes, values, and skills for active participation in society
7. Safety & Security	<ul style="list-style-type: none"> a. Living environment b. Supportive and informed primary caregiver(s) c. Caregiver's style for meeting the young persons needs for: <ul style="list-style-type: none"> i. safety ii. physical and emotional care iii. control, clear limits, boundaries iv. intellectual stimulation v. age-appropriate autonomy and responsibility

A category of “General Wellbeing” was added to those of the HCN well-being framework, which included quality of life and client satisfaction. This resulted in a total of eight domains to guide the search to identify measures. Ultimately, a measure or measures, that covered more than a single domain was to be preferred. However, the initial search was for measures relevant for assessing outcomes in any of these areas. To be suitable for routine use outcome measures need to take minimal time to complete. So again, the fewer measures required to assess outcome across domains, the better.

Criteria for the ideal measure

The tender document for this project outlined the criteria HCN required the final measures to meet. These are summarised below.

Feasibility: Is useable and useful for frontline staff

A feasible measure for HCN was one that was quick to use, could be used by a range of practitioners, and required minimal resources (time, training, materials, equipment etc.).

Can accommodate missing data

HCN said they often, for various reasons, could not gather complete data for clients. A measure that relied on all data being available at once, or that could not accommodate missing questions or forms, was not useful.

Scores/results are easily interpretable

HCN sought a measure with results that were easily understandable without advanced training. Many practitioners had little or no training in psychometric interpretation, so for a measure to be feasible it needed to be able to be used in this context.

Quantifiable constructs and results

The search was limited to measures that relied primarily on ratings, scales, or scores. Measures that may be useful, but relied on data such as comments, descriptions, examples, or anecdotes, were not to be considered.

Able to be repeatedly administered

HCN specifically asked for measures that could be used at frequent and possibly irregular intervals to suit the practical reality of working with their clients. They

specified they not be limited to pre- and post-test only, but be capable of giving a picture of progress during the intervention. A measure with good sensitivity to change and usefulness for ongoing progress monitoring was likely to fulfil this criterion.

Has appropriate norms

As the measures were intended for use across New Zealand, applicability to our local context was important. Normative data appropriate to the age group of HCN clients and the nature of their difficulties were also preferred.

Is “culture fair”

HCN defined this as using predominantly objective data to reduce reactivity to contextual factors. They felt that items relying on subjective ratings (e.g., based on opinion or impression) could cause bias depending on the orientation and cultural influences of the practitioner completing the form. Therefore, relying on factual, objective evidence was seen to reduce the effect of any bias present.

Has been tested for reliability and validity

A good measure was psychometrically robust; i.e., had evidence of reliability and validity. It was important to also have adequate face validity to help secure willingness and participation from both clients (and their family/whānau) and practitioners.

Can be used at individual and aggregate levels

HCN asked that measures useful for providing information about the progress of an individual client also provided information that would be useful at a wider level. The ability to aggregate data to indicate trends at a national level or for specific subgroups (such as ethnicity or gender) was important. This criterion eliminated from consideration any tests designed exclusively for group administration. In addition, it implied a single measure or standard set of measures was required because the same data type for each client was necessary in order to aggregate results.

The search for measures

Identifying potential measures

With the above parameters in mind, a search to locate potential measures was conducted. Locating measures for review and selection started with a search of electronic journal databases, test catalogues, the *Mental Measurements Yearbook*, and

previously published reviews of measures. The names of measures found were compiled into a single list and then narrowed down using a series of “screens”. The screening process compared the measures on the list with specific exclusion criteria, thereby decreasing the number of possible measures for review. The resulting shorter list of measures was used to begin more in-depth reviewing. Each of the steps in this process is outlined next.

Search of electronic sources

A list of keywords for entry into electronic journal databases was drafted. Because HCN cases are so varied, keywords concerning domain areas, problems, or disorders were not used. The rationale was to have the least restrictive initial search criteria to avoid missing potentially useful measures at the early stage of the search. Consequently, two lists of keywords were used: one to refine the search to measurement tools and one to narrow the age range (see Table 7).

Table 7 *Keywords Used to Search Electronic Databases*

<i>Keywords Related to Measurement Tools</i>		<i>Keywords Related to Age Group</i>	
Assess	Assessment	Adolescen*	Baby
Checklist	Diagnostic	Child*	Infant
Evaluation	Form	Kindergarten	Kindy
Indicator	Instrument	Preadolescent	Prepubertal
Interview	Inventor*	Prepubescent	Preschool
Measure*	Outcome	Preteen	Puberty
Outcomes	Profile	School	Teen
Psychometric*	Questionnaire	Teenage*	Toddler
Rating	Record	Young	Youth
Scale	Scales		
Schedule	Screen		
Self report	Survey		
Test			

* represents a search for words with multiple possible endings

These keywords were entered into several journal databases to retrieve articles likely to contain the names of measures that could then be researched further. The databases used were *PsychInfo*, *Medline*, and *Social Services Abstracts (SSA)*. In addition, the *Mental Measurements Yearbook Online (MMYBO)* was also searched. Some limiters were applied where available, which included: Human, English language, Post-1960, Not book reviews, Not dissertations. HCN had advised that the clients they funded were generally not youth offenders as the Ministry of Justice is responsible for

those young people, so two additional limiters were applied when possible: Not delinquent* and Not offender*.

Each search yielded a large number of results, almost 96000 in total (PsychInfo, 10238; Medline, 79537; SSA, 4130; MMYBO, 1921). The titles and abstracts for all these results were examined for the names of measurement tools or the suggestion that a relevant tool may be mentioned within the article. This process resulted in a list of measures that numbered almost 1600.

Search of catalogues

The second source used to locate possible measurement tools were library and retail catalogues. The Massey University, School of Psychology Test Library catalogue added almost 250 measures to the list. The New Zealand Council for Educational Research and the Psychological Assessment Resources catalogues were consulted, but only the 58 measures that were not already found elsewhere were added.

Search of published reviews

A number of reviews of measures have been undertaken in the past. Generally, these reviews were centred on a specific area of difficulty, professional orientation, or age group. As the present search was broader in its focus, measures identified by previous reviews were added, including those not short-listed in the review. The reviews selected were mainly conducted for New Zealand or Australia to increase the likelihood of finding locally relevant measures. A total of 356 measures were contributed from the following reviews:

- *Practitioner's guide to developmental and psychological testing* (Aylward, 1994)
- *Consumer measurement systems and child and adolescent mental health* (Bickman, et al., 1998)
- *The definition, diagnosis and treatment of children and youth with severe behaviour difficulties: A review of the research* (Church, 2003)
- *A review of progress and outcome measures: Use with Sensitive Claims clients in Aotearoa/New Zealand* (Fitzgerald, et al., 2006)
- *Child and youth outcome measures: Examining current use and acceptability of measures in mental health services and recommending future directions* (Merry, et al., 2004)
- *A comparison of commonly used treatment measures* (Zimmerman, 1996)

Screening the search results

The names of the measures located in the search were collated into a single master list. Once obvious duplications were removed, 1665 measures remained, although not all appeared relevant to HCN clients. To shorten the list to a workable size for reviewing, a series of minimally, but progressively, exclusive screens were applied using the HCN domains and criteria as guides.

The first screen

The master list was examined independently by two members of the research team, who classified every measure as either: “Keep”, “Discard”, or “Unsure”. In keeping with a conservative approach, a decision-making process based on the exclusion of unsuitable measures was adopted. The first screen was based on relevance for HCN purposes. This excluded measures that assessed static factors because, by definition, these could not be outcomes. In addition, specialist areas and areas not relevant to the eight domains were eliminated. Discarded measures included:

- Personality tests
- Projective tests
- Specialist neuropsychological tests
- Treatment fidelity/satisfaction only
- Measures designed for a specific population (e.g., hearing-impaired children)
- Measures related to a specific psychotherapy
- Measures labelled as “Research edition” or “Experimental edition”
- Measures related to a specific theoretical construct (e.g., defence mechanisms)

Measures classified as “Unsure” were discussed jointly in order to make a final decision. This classification applied when,

- More information was needed in order to make a decision
- Measures were consistent with the domains, but might not be useful
- Measures were not consistent with the domains, but might be useful

The first screen resulted in initial agreement between the researchers of 66% (Agreement: Keep, 690 measures; Discard, 383; and Unsure, 28; Total, 1101). Further information was found for the 28 measures that both were unsure about and the 564 measures disagreed on so that final classifications could be made. As a result, joint decisions were made to retain a further 200 measures for the next screen.

Before finalising the results of the first screen, a decision was made to discard measures applicable exclusively to children under five years old. Despite the HCN client age range including pre-school children, a conversation with HCN revealed they very rarely actually funded this age group. Consequently, a further 26 measures were discarded. In summary, of the original 1665 measures only 864 remained (51.9%) after the first screen.

The second screen

As 864 measures was still too large a list for reviewing, a second screen was undertaken. The alphabetically ordered list was divided at the halfway point, with each of the two researchers responsible for screening half the measures. Measures meeting the following exclusion criteria were discarded:

- Applicable *only* to those aged under 5 years
- Applicable *only* to those aged over 17 years
- Takes more than one hour to administer
- Is out of print or obsolete (e.g., based on DSM-III)
- Only the developer has published about the measure

Following the second screen, only 106 measures were retained (12.3% of the first screen; 6.4% of the master list).

The third screen

One final screen was applied to reduce the list further and to eliminate those not meeting the HCN criteria. The exclusion criteria for the third screen were:

- Takes more than 20 minutes to administer
- Requires advanced qualifications to use
- Not available for use in New Zealand

It became clear that the list consisted of two types of measures, (1) those with an in-depth focus on a narrow area of concern (e.g., spelling, depression, autism) and, (2) those that were more wide-ranging in their coverage. Taking into account feasibility for HCN teams, the narrow/in-depth measures were discarded. Because the search was to find and recommend as few measures as possible, those having a broad focus on a number of dimensions were preferred.

With 91 measures being discarded in the third screen, 15 measures remained. Consideration of these measures in terms of the HCN domains of well-being showed

none had been identified that tapped the “Cultural Identity” domain. Two additional measures were added after cultural consultation, giving a final 17 measures for review (see Table 8).

Table 8 *Search Results: Measures Remaining After the Screening Process*

<i>Area of coverage (n)</i>	<i>Name of measure (Abbreviation)</i>
Culture (1):	Multigroup Ethnic Identity Measure (MEIM)*
Education (1):	Canadian Achievement Tests (CAT3)
Family (5):	Family Environment Scale (FES) Index of Family Relations (IFR) McMaster Family Assessment Device (FAD) North Carolina Family Assessment Scale (NCFAS2.0) Self-Report Family Inventory (SFI)
Mental Health (3):	Hua Oranga* Pediatric Symptom Checklist (PSC) Strengths and Difficulties Questionnaire (SDQ)
Physical Health (1):	Child Health and Illness Profile (CHIP)
Several Areas of Focus (6):	Behavior Evaluation Scale (BES3) Child Health Questionnaire (CHQ) Clinical Assessment Package for Risks & Strengths (CASPARS) Health of the Nation Outcome Scales: Child & Adolescent (HoNOS-CA) Ohio Youth Problems, Functioning, and Satisfaction Scales (Ohio Scales) Youth Outcome Questionnaire (YOQ)
Total:	17

* Added to the final list following cultural consultation

The measures in the final list did not require advanced qualifications, took less than 20 minutes to complete, were available for use in New Zealand, had research data available, and were suitable for 5-18 year olds. Following the reduction of the initial search results to these 17 measures, reviews were conducted with the intention to decide which, if any, would be appropriate for HCN's purposes.

Brief review of measures

More information regarding the final 17 measures was sought and each was reviewed. Some measures received favourable reviews, some were adequate measures but not suited to HCN, and a small number were deemed unsuitable. A short description of each measure follows, with a brief summary of the review it received.

Measures of culture

1. Multigroup Ethnic Identity Measure

Description

Rather than the unwieldy scenario of multiple measures covering all cultural/ethnic groups HCN teams work with, the *Multigroup Ethnic Identity Measure* (MEIM, Phinney, 1992) provides a more pragmatic approach. It is a single measure that focuses on those common elements of ethnic identity that are found across groups. The MEIM was designed for adolescents and young adults and is a pen-and-paper, self-report measure consisting of 12 items, rated from “1 strongly disagree” to “4 strongly agree”. Examples of items are, “I have a clear sense of my ethnic background and what it means for me”, and “I participate in cultural practices of my own group, such as special food, music or customs”. The MEIM is available for free download from the developer’s website, <http://www.calstatela.edu/academic/psych/html/download.htm>.

There are two scales, Ethnic Identity Search (5 items), and Affirmation, Belonging, and Commitment (7 items), and an overall score can be calculated. With no reverse scoring, scores are the mean of scale items (or all items for the overall score). This provides a means for managing missing items. The scales have been supported by factor analysis (Roberts, et al., 1999) and the measure has good reliability (Phinney, 1992). The scale structure has been undergoing revision to improve the theoretical strength of the measure and the psychometric profile.

Suitability for HCN

Unfortunately, the MEIM has only been used in a small number of research studies with high school students and undergraduate university students in the USA, so it could still be regarded as an experimental measure. The language and level of cognitive sophistication required to understand the concepts mean it is unsuitable for use with younger children. However, possible future development of the measure could include a parallel version for a parent/caregiver to complete for their young child, or a simpler child self-report.

There is no data available for how this measure performs for those with high and complex needs, or with populations outside North America. No information was found regarding its use as a measure of outcome, although the item phrasing might be appropriate for noting change following interventions to increase cultural identity or

sense of affiliation. Further research on test-retest reliability and sensitivity to change is required before the MEIM could be considered for use in this manner.

Measures of education

2. Canadian Achievement Tests – Third edition

Description

The Canadian Achievement Tests (CAT-3; publisher's website, <http://www.canadiantestcentre.com/index.htm>) is a battery designed to assess students' educational progress. It was developed to reflect the Canadian school curricula, with themes reflecting their societal diversity. There are a range of tests available including, Word Analysis, Vocabulary, Spelling, Language/Writing Conventions, Computation, Dictation, Mathematics, and Writing. The basic battery, however, consists only of a combined reading/language test and a mathematics test, and it was this version that was considered for HCN purposes. Depending on the age and level of the child, each test in the basic battery takes approximately 20-40 minutes to administer. This is outside the range specified in the measure screening process, but as there were no other appropriate educational measures, the CAT3 was retained for review.

Suitability for HCN

The CAT3 was normed in the year 2000 with 44000 Canadian 6-18 year olds. It is an expensive measure and only available for purchase by educational institutions or registered homeschoolers, which limits its feasibility with HCN. In addition, the suitability of the tests and norms to New Zealand is questionable. Because the CAT3 was developed to align closely with the Canadian curricula, it might include non-New Zealand curriculum material, not include important New Zealand material, or differ in terms of the age at which similar content is taught here. In that case, New Zealand students would be disadvantaged with a test such as this. Finally, the very lengthy administration time would be highly impractical for HCN teams, particularly if used as a standard routine measure with all clients.

Measures of family

3. Family Environment Scale

Description

The Family Environment Scale (FES; Moos & Moos, 1994) was developed to measure the environmental characteristics of families. It is based on a three dimensional conceptualisation of the family with 10 sub-domains within those dimensions. These dimensions (sub-domains) are, (1) Relationships (cohesion, expressiveness, conflict), (2) Personal Growth (independence, achievement orientation, intellectual-cultural orientation, active-recreational orientation, moral-religious emphasis), and (3) System Maintenance (organisation, control). This self-report measure can be completed by any family member over 11 years old, facilitating a calculation of the agreement between individuals. The 90 true/false items take 15-20 minutes to complete. Three different forms can be administered, “Real” (actual perceptions), “Ideal” (ideal family environment), and “Expectations” (expectations about family settings).

Ten scores are derived to give an overall profile of family environment and used to categorise the family as one of three environment types. Clinical cut-off scores are also provided. The normative samples are 1125 “normal” and 500 “distressed” families, with the normal families drawn from a variety of geographic, ethnic, family type and age groupings.

Suitability for HCN

The developers reported internal consistency for the FES that varied across the sub-scales and five studies (ranging from N = 356 to N = 904) from fair (.58) to excellent (.79; Moos, 1990). In addition, these were all lower than those reported for their previous normative sample (Moos, 1990). A large set of normative data was later collected in Australia for 1289 adolescents aged 11-18 years (Boyd, Gullone, Needleman, & Burt, 1997). Boyd and colleagues added to previous criticisms of the measure’s reliability, finding inadequate estimates for their large sample. In addition, Bickman and colleagues’ (1998) comprehensive review of outcome measures did not recommend the FES as they highlighted concerns about reliability, validity and the evidence of factor analytic studies.

The FES can be purchased without any particular qualification requirements (from <http://www.mindgarden.com>) and has been translated into five languages for research

purposes. However, no evidence of its use with extended or indigenous families could be found. Unlike many other measures, the FES has not been reviewed or updated since its development. Due to concerns about the length of time to administer this measure to multiple family members, psychometric problems, and the relative expense involved with its use, the FES was not deemed suitable for use by HCN.

4. Index of Family Relations

Description

The Index of Family Relations (IFR; Hudson, 1990a) is a brief screening device for intrafamilial stress. It is a self-report measure for children over 12 years old about how they feel about their family as a whole. Administration time is estimated to be 10 minutes at most (Hudson, 1982). Designed to assess the severity of family relationship problems, its 25 items are rated on a 7-point scale according to the frequency with which the problem occurs. Items are balanced between positive and negative views of families, such as “There is too much hatred in my family”, and “I feel proud of my family”. Reverse scoring applies to a number of items, which makes its use more complicated for novice test users, and a total score between 0 and 100 is calculated. The clinical cut-off is 30, with greater amounts of family difficulties indicated by higher scores (Hudson, 1990b).

Suitability for HCN

Because the measure can only be completed by children with a reading age of at least 10 years, difficulties with missing data could arise for children who do not read well enough to answer all items or inadvertently miss questions, and there is no parallel adult form to help mitigate this problem. The publishers (<http://www.walmyr.com>) reported reliability of at least .90 and good validity. No test-retest data were available, nor has the factor structure or effect of socially desirable responding been investigated.

There are no restrictions on who can administer the IFR and it was reasonably priced. The IFR was potentially useful for HCN, however more psychometric evidence from independent sources was required before its use as an outcome measure or tracking device could be recommended.

5. McMaster Family Assessment Device – Version 3

Description

The McMaster Family Assessment Device (FAD; Epstein, Baldwin, & Bishop, 1983) is a short, self-report assessment of families' emotional relationships and functioning. This 60-item measure is based on the McMaster Model of Family Functioning and the dimensions tapped reflect this. Six of the scales are based on these dimensions: Problem Solving, Communication, Roles, Affective Responsiveness, Affective Involvement, and Behaviour Control. In addition, there is a General Functioning scale, which gives an overall assessment of the family. Each scale contains items related to both healthy and unhealthy functioning and these are rated using a 4-point scale ("strongly agree" to "strongly disagree"). The hand scoring procedure requires some ratings to be transformed before averaging the item scores to give scale scores between 1 (healthy) and 4 (unhealthy).

A useful addition to the FAD is two pages for capturing clinically useful information about the family. This includes demographic details, marital status and history, family income, and current employment. Ethnicity/race questions would need to be reworked for use in New Zealand to reflect appropriately the diversity of cultures here.

Norms from psychiatric, medical, and non-clinical samples are available. Internal consistency ranged from good to excellent for the seven scales, and one week test-retest reliability from .66 (Problem Solving) to .76 (Affective Responsiveness; Miller, Epstein, Bishop, & Keitner, 1985). Construct validity has been supported with over 90% of items loaded on factors as hypothesised (Kabakoff, Miller, Bishop, Epstein, & Keitner, 1990), although subsequent research has not replicated this completely (Ridenour, Daley, & Reich, 1999). Concurrent validity has been established with other well known assessments of families (Epstein, et al., 1983; Miller, et al., 1985). Miller and colleagues (1985) also found low correlations of the FAD with social desirability.

Suitability for HCN

Psychometric support for the FAD is good and the link to an underlying conceptualisation of family functioning is a strength. It is in the public domain and requires no particular level of qualification to administer. The measure takes approximately 15-20 minutes to complete, which might be too long for HCN if measures of other domains are also being used. Because the FAD can only be

completed by those over 12 years old, its usefulness with younger clients might be further limited.

6. North Carolina Family Assessment Scale – Version 2.0

Description

The North Carolina Family Assessment Scale (NCFAS; Kirk & Reed-Ashcraft, 1998b) was designed to measure aspects of family functioning using the ratings of the professional working with the family (most often a Social Worker). It is a practice-based measure that was implemented state-wide in the Intensive Family Preservation Services in North Carolina, USA in the 1990s (Kirk & Reed-Ashcraft, n.d.-b). The NCFAS is based upon an ecological model of families and measures functioning with five subscales: Environment, Parental Capabilities, Family Interactions, Family Safety, and Child Wellbeing. The 36 items are rated on a scale from +2 (“Clear Strength”), to 0 (“Baseline/Adequate”), to -3 (“Serious Problem”). Space is provided for rating at intake and closure. The professional completing the NCFAS rates each item based on their observations and information gathered about a family during the course of their work with them. A separate “Definitions” booklet provides anchors for the extreme and midpoint ratings of each item. For example, on the Family Interactions subscale, item 2, “Bonding with children” reads (Kirk & Reed-Ashcraft, 1998a, p. 7),

<i>+2 Clear Strength</i>	Refers to caregivers’ healthy closeness with their child and their ability to nurture a child. Caretakers encourage appropriate independence for child, and give love and attention freely to child. They respond to child’s needs appropriately, and have a sense of attachment to child.
<i>0 Baseline/Adequate</i>	Refers to caregivers’ ability to be close to their child. Caregivers do not openly encourage independence for their child, and may not give affection openly to child. However, child’s needs appear to be met.
<i>-3 Serious Problem</i>	Refers to caregivers’ inability to form a close relationship with their child, and inability to nurture their child. Caretakers are resentful, rejecting, or detached from their child. Also refers to caregivers’ non-responsiveness, inappropriate responsiveness, or extreme enmeshment with their child.

Suitability for HCN

Reliability evidence showed alpha coefficients for the subscales in the excellent range for intake and closure (Kirk & Reed-Ashcraft, n.d.-a). No test-retest data could be located, although there was evidence of sensitivity to change between intake and closure (Kirk & Reed-Ashcraft, n.d.-a). Validity studies suggested support for the

subscales, although only three of five factors identified were supported. In a report by the Bay Area Social Services Consortium in California, USA (Johnson, et al., 2006) the NCFAS was selected as the most useful for child welfare settings from seven family assessment instruments evaluated, due to its strengths orientation and extensive testing with that population.

The NCFAS takes approximately 15 minutes for a professional to complete. There is no particular qualification level needed to use the measure and it appeared to be in the public domain (download from <http://www.nfcpn.org>). The issues covered and range of items appeared highly relevant for HCN. The NCFAS needed to be supplemented with data from other measures in order to have broad coverage of outcome domains.

7. Self-report Family Inventory

Description

The Self-report Family Inventory (SFI) describes family members' perceptions in the six domains of the Beavers' Systems Model of Family Functioning (Beavers & Hampson, 2000): Family Health/Competence, Conflict, Cohesion, Expressiveness, and Directive Leadership (Beavers, Hampson, & Hulgus, 1985). The 36 items are rated on a Likert-type scale and take around 15 minutes to complete. Examples of SFI items are, "Household members pay attention to each other's feelings", "Household members put each other down", and "Our household is happy most of the time". The measure is completed by any family members over 11 years old, which provides an opportunity for more than one source of data and comparison of different family members' perspectives.

Normative data is for competent non-clinic families, less competent non-clinic families, and clinic families. One and three month test-retest reliability ranged from fair (Directive Leadership) to excellent (Expressiveness), and internal consistency was slightly stronger at .86 (Hampson, Hulgus, & Beavers, 1991). Factor analysis supported construct validity (Beavers, et al., 1985), and convergent and concurrent validity have been demonstrated with other measures of family functioning (Beavers, et al., 1985; Epstein, et al., 1983; Hampson, et al., 1991).

Suitability for HCN

One strength of the SFI is that it is based on a theoretical model of family functioning, with factor analysis supporting this. While internal consistency was good, more work

needed to be done to improve test-retest reliability. While this measure showed promise, there were doubts about its usefulness for measuring individual outcomes for a young person. Rather, it appeared better suited for interventions targeting a whole family unit.

Measures of mental health

8. Hua Oranga

Description

Hua Oranga is a 16-item outcome measure specifically developed for use with Māori (Kingi, unpublished; cited in S. Bennett, personal communication, June 2006). It is based on a holistic model of health, Te Whare Tapa Wha (Durie, 2001), and asks respondents to rate four items related to each of the dimensions of the model: Te Taha Wairua, Taha Hinengaro, Te Taha Tinana, and Te Taha Whānau⁶. There are three forms for completion by tāngata whaiora⁷, whānau, and clinician. Administration, scoring, and interpretation are straightforward.

Suitability for HCN

Hua Oranga was a relatively new measure and was being validated at the time of this review (S. Bennett, personal communication, June 2006). Anecdotal feedback suggested excellent face validity and acceptance by Māori, but unfortunately, the results of practitioner consultation suggested administration with those less than 16 years old was difficult, due to poor understanding of some questions (S. Bennett, personal communication, August 2006). However, the clinician and whānau forms could still be used in these situations.

A project was underway, using a re-worded version of Hua Oranga in order to make it more suited to repeat administration (S. Bennett, personal communication, August 2006). Original items were reworked, such as, “As a result of the intervention at this point does the client feel healthier from a spiritual point of view?”, which was rated on a 5-point scale from “much more” to “much less”. The new item reads, “In your opinion, at this point in time does the client feel healthy from a spiritual point of view?” and is rated from “strongly agree” to “strongly disagree”. This change would increase the

⁶ The dimension of spirituality, the dimension of the mind, the dimension of the body, and the dimension of the family.

⁷ Client/mental health consumer

usefulness of the measure for tracking progress with multiple administrations over time. However, this type of question, although providing quantifiable data, is subjective and what HCN preferred to avoid. In addition, Merry and colleagues (2004) reported that young people were unsure how to interpret this question about spirituality.

Although not suitable for use with all HCN clients, Hua Oranga is the only New Zealand developed measure based on a Māori view of health. No normative data were available, but may be as it is further developed. As mentioned above, the measure was not appropriate for children, but if this changed in the future, its usefulness as an outcome measure should be reassessed.

9. Pediatric Symptom Checklist

Description

The Pediatric Symptom Checklist (PSC; Jellinek, et al., 1999) is a screening tool to alert practitioners to difficulties that may indicate current, or lead to future, psychosocial problems. The PSC has been designed for administration in primary care settings, such as a General Practitioner's waiting room. Consequently, it is deemed to have high face validity, require little or no introduction, and has clear instructions regarding completion. Thirty-five items are rated by a parent/caregiver according to how true the statements are for the young person: "0 Never", "1 Sometimes", or "2 Always". A range of behaviours and emotional problems are described, for example, "Daydreams too much" and "Does not listen to rules". The PSC takes less than five minutes to complete.

A total score ranging from 0-70 is calculated by summing the items, and scores above the clinical cut-off indicate significant psychosocial impairment. For children aged 4-5 years the total score is based on 31 items (clinical cut-off: 24), and all items are used for young people aged 6-16 years (clinical cut-off: 28). Three factor scores can also be determined for young people whose total scores are in the clinical range, which are Attention, Externalising, and Internalising. A parallel youth self-report version has been developed for use with young people aged 9-14 years (PSC-Y; Pagano, Cassidy, Little, Murphy, & Jellinek, 2000).

Suitability for HCN

Both the parent and youth forms have good psychometric properties (Pagano, et al., 2000) and are freely available for use. In addition to the original English language

version, four other translations exist. As the PSC and PSC-Y enable multi-informant administration, are quick to use, and have a well-established body of research behind them, they were deemed potential measures for use by HCN teams.

10. Strengths and Difficulties Questionnaire

Description

The SDQ (Goodman, 1997) was developed in the UK as a brief screening device. There are three independent sections to the SDQ, although not all need to be included to use the measure. The first, “Psychological Attributes”, forms the core of the measure and consists of 25 items, such as “Has at least one good friend” and, “Often lies or cheats”. There are five items for each of the following areas: emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and prosocial behaviour. The items are rated as either “Not true”, “Somewhat true”, or “Certainly true” for the young person over the last six months. The prosocial behaviour items add to a total Strengths score, with items in the remaining areas contributing to a total Difficulties score.

The second section is an “Impact Supplement”, which gathers detailed information about problems the respondent identified as occurring in the last six months. Questions cover chronicity, degree of distress, interference with everyday social life, and burden on others. The 4-point scale (“Not at all”, “A little”, “A medium amount”, and “A great deal”) provide useful information for practitioners.

The third section, “Follow Up” adds two questions regarding the respondent’s view of the intervention. First, a rating of relative change in the problem (5-point scale: “Much worse” to “Much better”), and second, about other effects the intervention had. These questions are answered for the last month.

The SDQ is available in almost 60 languages in addition to three English language versions (Australia, UK, and USA). Norms have been developed for a number of countries including, Britain, Finland, Germany, Sweden, and the USA (see www.sdqinfo.com). The SDQ has parallel forms available for parents of 4-10 and 11-17 year olds, teachers of 4-10 and 11-17 year olds, and self-report forms for 11-17 year olds. Using multiple informants was found to increase the sensitivity of the measure in screening for mental health difficulties (Goodman, Ford, Simmons, Gatward, & Meltzer, 2000). Evidence suggested this measure has a sound factor structure and good

psychometric properties (Goodman, 2001), and support for its validity (Goodman, Meltzer, & Bailey, 2003).

Suitability for HCN

The focus on positive aspects is important to track, as there might be little change in difficult behaviour but an increase in prosocial behaviour following an intervention, particularly with complex cases. The SDQ can be downloaded for no cost with all forms, instructions, and transparency scoring templates available. Clinicians in New Zealand are familiar with this measure and appreciate its ease of use, quick administration, and applicability to practice (Merry, et al., 2004). However, it gives a broad picture of strengths and difficulties over the last six months, which might not provide the sensitivity to detect change in a timely enough manner for clinicians to use. Although it was designed as a screening tool rather than as an outcome measurement tool, with the follow-up section, it could be useful.

Measures of physical health

11. Child Health and Illness Profile

Description

The Child Health and Illness Profile (CHIP; Starfield, et al., 1993), has two editions; one for children 6-11 years old (CHIP-CE), and one for adolescents 11-18 years old (CHIP-AE). There is no specific training required for purchasers of these measures, but they are expensive (see www.chiphealthprofile.org). The two editions have adequate psychometrics, but differ slightly so are described here separately.

The CHIP-CE has forms for both parent and child completion. Both forms cover signs of wellbeing and illness, health-related and problem behaviours, school performance, and involvement with family and peers. The child form requires a reading age of 8 years, and includes illustrations for each of its 45 items assessing four domains: satisfaction, discomfort, resilience, risk avoidance, and achievement. The parent form has 41 additional optional items that assess disorders and medical conditions. Scores in each domain are reported as T-scores, for comparison with “average” health profiles of children.

The CHIP-AE is used with young people over 11 years old and a reading age of at least 10 years. Six health domains are covered: satisfaction with health, discomfort,

achievement, risks, resilience, and disorders; 108 items and 46 disease/injury specific questions in total. The period used for the 5-point rating scale varies across items from 28 days to 12 months. Domain scores are the mean of item ratings, allowing for interpretation when up to 30% of items are missed. These raw scores are converted to T-scores scores based on a USA normative sample.

Suitability for HCN

While this was the only physical health measure considered, it was not judged suitable for HCN for a number of reasons. The lengthy administration time (estimates range from 20-45 minutes) made feasibility for busy clinicians doubtful. While it might appear useful on an individual basis, it was actually developed to collect health data about populations; consequently may be unsuitable for use as an outcome measure for individuals. Finally, this was an expensive measure and while this alone should not eliminate appropriate measures from consideration, it further added to its unsuitability.

Measures with several areas of focus

12. Behaviour Evaluation Scale – 3rd edition

Description

The Behaviour Evaluation Scale (BES-3) is commercially published, so there is a moderate cost involved in its use (see <http://www.hes-inc.com>). It was developed to assist schools' decision-making about placement and programming for students in accordance with the definitions of emotional disturbance and behavioural disorders from the USA *Individuals with Disabilities Education Act*. In particular, it was intended for use with those who have "handicapping conditions", such as learning disabilities, mental retardation, or physical impairments. Factor analysis created five subscales: learning problems, interpersonal difficulties, inappropriate behaviour, unhappiness/depression, and physical symptoms/fears. The BES-3 is available in Short and Long forms, which both have a School (54 or 73 items) and Home version (52 or 76 items). Completion time estimates are 15 and 20 minutes each for the Short and Long versions, respectively. The USA standardisation sample was large (5124 for the school version and 4643 for the home version) with an age range of 4-19 years. According to the publishers, reliability and validity are adequate to good.

Suitability for HCN

While the scale was useful for defining the presenting problems, it was the accompanying “Intervention Manual” that appeared the most useful aspect of the kit. It contained a large number and wide range of intervention strategies for inclusion in a plan to address issues identified by each item of the measure. Thus, the real strength was not its comprehensiveness or psychometric properties, but the link it had to the development of individualised intervention plans.

In terms of usefulness for HCN teams, the BES-3 had potential to assist with assessment and intervention planning, but as an outcome measure was limited. In addition, the focus was educational and even within that domain, had specific and rather restricted utility.

13. Child Health Questionnaire

Description

The Child Health Questionnaire (CHQ; www.healthact.com) is a generic quality of life measure for use with children aged 5-18 years. The CHQ has two parent forms (a 50-item and a 28-item version) and a youth self-report form for those aged at least 10 years (87 items). It has been translated into multiple languages in addition to four English language versions (USA, UK, Canada, and Australia). The 13 areas assessed are, physical functioning, role/social emotional-behavioural, role/social physical, bodily pain, general behaviour, mental health, self-esteem, general health perceptions, change in health, emotional and time impact on parents, family activities, and family cohesion.

The parent-report 28-item version has comparable psychometric properties to its longer counterpart, except for the reliability of the subscales. For this reason, the shorter version is recommended only for situations where a summary score will suffice (Raat, Botterweck, Landgraf, Hoogeveen, & Essink-Bot, 2005). The longer youth-report form is reliable and valid (Waters, Salmon, Wake, Wright, & Hesketh, 2001). The parent form has been evaluated for use in Australia where the strong psychometrics previously published were replicated, but the factor structure differed, as did scores on some subscales (Waters, Salmon, & Wake, 2000). The CHQ is a commercially published measure so is quite expensive to use.

Suitability for HCN

Although the CHQ claimed to measure physical and psychosocial health areas, it was weighted toward physical health. This impression was supported by the fact that much of the published research validating the CHQ was conducted with young people experiencing chronic medical problems, for example, diabetes (Cameron, Smidtst, Hesketh, Wake, & Northam, 2003) and cancer (Sung, et al., 2003). Also, factor analytic findings indicated quite different factor structures for physically healthy respondents and those with chronic medical conditions (Drotar, Schwartz, Palermo, & Burant, 2006). Therefore, the CHQ was not considered suitable for HCN, unless evaluating health status was indicated for a particular client.

14. Clinical Assessment Package for Risks and Strengths***Description***

Developed by Professor Gilgun of the School of Social Work, University of Minnesota, the Clinical Assessment Package for Risks and Strengths (CASPARS; Gilgun, 1999) is a battery of five individual measures. They are family's and child's embeddedness in community (13 items), emotional expressiveness (8 items), current relationships among family members (21 items), peer relationships (17 items), and sexuality (10 items). Clinicians highly rated the CASPARS for usefulness, ease of use and interpretation (Gilgun, 2004). The items are rated using a 7-point scale ranging from "asset" to "risk factor", with "not known/not observed/absent" as the midpoint. The ratings have anchors at the extremes, for example "family members have pleasant social relationships with extended family" (asset) and "family members do not have pleasant relationships with extended family members" (risk). The emphasis on both positives and negatives yields two total scores, consistent with clinicians' aim to both increase assets and decrease risk (Gilgun, 1999; Gilgun, Keskinen, Marti, & Rice, 1999). CASPARS can be downloaded at no cost (see http://ssw.che.umn.edu/Faculty_Profiles/Gilgun_Jane).

Suitability for HCN

CASPARS' psychometric properties were good, however the standardisation sample of 146 was small (5-13 years; Gilgun, 1999). Excellent reliability was reported (above .90) for all scales, and validity was acceptable (Gilgun, et al., 1999). The range of issues covered, while relevant to HCN clients, was limited. Rather than the whole package being used, individual measures might be useful for HCN when supplemented by measures in other domains.

15. Health of the Nation Outcome Scales for Children and Adolescents

Description

This measure assesses behaviour, symptoms, and functioning of children and adolescents with mental health difficulties. The Health of the Nation Outcome Scales for Children and Adolescents (HoNOS-CA; Gowers, Harrington, Whitton, Beevor, et al., 1999; Gowers, Harrington, Whitton, Lelliott, et al., 1999) was developed in the UK after extensive field tests. It has been mandated for routine use in Australian and New Zealand CAMHS, although the psychometrics were poor for use in routine practice here (Trauer, et al., 2004). Given its widespread use, New Zealand normative data may be collected in due course, which would be a real advantage for this measure.

The clinician completed HoNOS-CA is considered to be the primary measure, however, a youth self-report (Gowers, et al., 2002) has been developed. There are 13 items, eight of which are clearly mental health focussed and they are rated on a 5-point scale from “Not at all” to “Severely” a problem. The items are difficulties with,

1. disruptive, aggressive, or anti-social behaviour
2. over-activity, attention, and concentration
3. non-accidental self injury
4. alcohol, substance/solvent misuse
5. scholastic or language skills
6. physical illness or disability problems
7. hallucinations or delusions
8. non-organic somatic symptoms
9. emotional and related symptoms
10. peer relationships
11. self-care and independence
12. family life and relationships
13. poor school attendance

The scale is accompanied by a booklet of definitions for rating each item, which guides the practitioner in assessing which problems to rate where. A total scale and five subscale scores (Behaviour, Impairment, Symptoms, Social, and Information) are computed by simply summing the appropriate item ratings.

Suitability for HCN

The informant forms are parallel in content, but this is problematic for HoNOS-CA. For example, one item on the youth form reads, “Have you experienced difficulties keeping

up with your usual educational abilities?” First, if the answer is “Severely”, then the client might not read well enough to answer the question. Second, the inclusion of the word “usual” makes the question difficult to interpret. That is, if the young person started with severe educational difficulties and still has severe educational difficulties, then the correct answer should be “Not at all”. Unfortunately, a number of items seem to be similarly flawed. Another problem on both forms is the use of double-barrelled questions such as, “Have you suffered from self-induced vomiting, head/stomach aches with no physical cause, bedwetting, or soiling?” Again, this poses problems for accurate rating, interpretation and making clinical use of the data.

The psychometric properties of the HoNOS-CA, as reported by the developers, are generally adequate, face validity is considered high, as is convergent validity and sensitivity to change (Garralda, Yates, & Higgins, 2000; Yates, Garralda, & Higgins, 1999; Yates, Kramer, & Garralda, 2006). There is also evidence supporting its concurrent validity with the parent-rated CBCL, self-report YSR, and teacher-rated TRF (Hanssen-Bauer, Langsrud, Kvernmo, & Heyerdahl, 2010). However, a study with New Zealand inpatient adolescents found that while it gave a good indication of global functioning at intake, it missed highly clinically relevant changes over the following six months (Harnett, Loxton, Sadler, Hides, & Baldwin, 2005), which presents a problem for its use as a routine measure of progress and outcome. While HoNOS-CA could have utility as a general mental health screen and despite its widespread use, the heavy focus on mental health problems and the structural problems identified indicated it was not the most useful outcome measure reviewed.

16. Ohio Youth Problems, Functioning & Satisfaction Scales

Description

The Ohio Scales (Ogles, et al., 1999; Ogles, Melendez, Davis, & Lunnen, 2000) is currently used by all publicly funded mental health services in Ohio, USA. It was purpose designed as a routine outcome measure for services with young people aged 5-18 years old. Three forms are available with most sections parallel on each. The Worker form differs from the Parent and Youth self-report (12-18 years) forms in that the Satisfaction (with the service and intervention) and Hopefulness scales are replaced with the *Restrictiveness of Living Environments Scale* (ROLES, a stand alone previously developed measure; Hawkins, Almeida, Fabry, & Reitz, 1992). The first part of all forms is the Problem Severity scale, with 20 items rated for frequency on a 6-point scale from “Not at all” to “All the time”, for example, “Getting into fights” and

“Using drugs or alcohol”. The last part of all forms is the Functioning scale and these 20 items are rated on a 5-point scale from “Extreme troubles” to “Doing very well”, for example, “Getting along with family” and “Participating in recreational activities”. There is no reverse scoring and ratings are summed to give a total for each scale. There are procedures for accommodating missing items, although if more than three are missing the scale is no longer valid.

Suitability for HCN

The forms are brief and take approximately 5-10 minutes to complete. A comprehensive manual is available and is sufficient training for people working with youth to use these scales, although training videos are available for no cost online (Ogles & Williams, 2009). A one-off license fee is associated with the measure’s use and the amount payable depends on the type of use required (e.g., single practitioner or nationwide organisation), but is modest.

The scales have good to excellent reliability, validity and sensitivity to change (Ogles, et al., 2001). One-week test-retest reliability was adequate, but better for the parent form (.67-.88) than the youth self-report (.43-.74). Initial research into concurrent and construct validity has been positive (Ogles, et al., 2000). The many years of state-wide use as a routine outcome measure means there is a huge clinical databank, which could serve as excellent normative data (in 2010, the sample size for each age band (e.g., 5-8 year olds) was between 62000 and 77000).

The mental health orientation of the wording on many items could be a concern for users working in other contexts. Although it does not cover preschool children, the breadth of content, ease of use, proven usefulness as a routine outcome measure, short administration time, and opportunities for triangulation of data from multiple informants, made the Ohio Scales an attractive potential outcome measure.

17. Youth Outcome Questionnaire

Description

The Youth Outcome Questionnaire (YOQ; Burlingame, et al., 1996) is the downward extension of a widely used adult outcome measure. It was purpose developed to be a brief measure of outcome, repeatedly administered on a routine basis, that was sensitive to change, and for little cost (Wells, Burlingame, Lambert, Hoag, & Hope,

1996). The developers intended its 64 items to take less than seven minutes to complete.

Parent and self-report forms address behaviours, situations, and moods common to “troubled teenagers”. The YOQ has subscales covering risk factors related to, (1) intrapersonal distress, (2) interpersonal relations, (3) somatic, (4) critical items, (5) social problems, and (6) behavioural dysfunction. The critical items subscale constitutes a partial risk evaluation, and high scores on these items indicate that a more detailed risk assessment should be completed. It also includes a brief prognosis questionnaire to determine the degree to which these risk factors might affect the change process. Three primary areas for prognosis are covered: history of mental illness in the immediate and extended family, current social environment (including integrity and stress on the family structure and socioeconomic status), and the client’s medical, developmental, and mental health history (Wells, et al., 1996). There are cut-off scores to provide normative comparison.

Suitability for HCN

The YOQ has excellent psychometric properties and high face validity (Burlingame, et al., 2001; Wells, et al., 1996). For the total score, internal consistency and test-retest coefficients are excellent (Ridge, Warren, Burlingame, Wells, & Tumblin, 2009; Whoolery, 1997). There is also evidence of good concurrent and divergent validity (Dunn, Burlingame, Walbridge, Smith, & Crum, 2005; Ridge, et al., 2009). The measure needs no specific qualification level and requires a one-off licence fee to use (\$75-\$3500, depending on the size of the organisation). It is a robust and well-supported measure, although the age range limited its use with HCN clients.

Summary

A wide search for outcome measures was narrowed down from 1665 to 17 measures using a three stage screening process. The final 17 measures were reviewed and evaluated for applicability with HCN clients. A summary of their ability to meet the requirements of a good routine outcome measure is presented in Table 9. It is important to note the limitations of the search. In particular, it did not include measures exclusively used with those aged under 5 or over 17 years, specialist measures requiring higher tertiary qualifications, or those covering only narrow areas of difficulty. Following the brief reviews, six measures were short-listed as being most suitable,

although no measure fit all criteria and could encompass the diversity of needs that HCN young people present. The six short-listed measures were,

- Clinical Assessment Package for Risks & Strengths
- North Carolina Family Assessment Scale
- Ohio Youth Problems, Functioning & Satisfaction Scales
- Pediatric Symptom Checklist
- Strengths & Difficulties Questionnaire
- Youth Outcome Questionnaire

The measures were considered by a group of clinical psychologist/researchers and presented for feedback from HCN practitioners before the final selection of a measure was made. This expert and practitioner consultation is described in the next chapter.

Table 9 Comparison of the 17 Reviewed Measures Against the Requirements of a “Good Routine Outcome Measure”

Name of Measure	Robust psychometrics	Sensitive to change	Items applicable to HCN clients	Covers symptoms, functioning, satisfaction & quality of life	Considers family, developmental, educational & social aspects	Multiple perspectives	Has cut-off scores or descriptive ranges, & normative data	Results useful clinically, & at other levels	Maximum 15 minutes to complete	Needs little training to use	Minimal or no cost	User-friendly	Short-list?
<i>Multigroup Ethnic Identity Measure</i>	Some issues	Unknown	Some issues	No	No	No	Some issues	Unknown	Yes	Yes	Free	Some issues	No
<i>Canadian Achievement Tests</i>	Yes	Unknown	Some issues	No	Some issues	No	Yes	Yes	No	No	No	Yes	No
<i>Family Environment Scale</i>	Some issues	Unknown	Yes	No	Some issues	Yes	Yes	Yes	No	Yes	No	Yes	No
<i>Index of Family Relations</i>	Some issues	Some issues	Yes	No	Some issues	No	Yes	Yes	Yes	Yes	Yes	Some issues	No
<i>McMaster Family Assessment Device</i>	Yes	Yes	Yes	No	Some issues	Yes	Yes	Yes	No	Yes	Free	Some issues	No
<i>North Carolina Family Assessment Scale</i>	Some issues	Yes	Yes	No	Some issues	No	Yes	Yes	Yes	Yes	Free	Some issues	Yes
<i>Self-report Family Inventory</i>	Some issues	Some issues	Yes	No	Some issues	Yes	Yes	No	Yes	Unknown	Unknown	Yes	No
<i>Hua Oranga</i>	Some issues	Unknown	Some issues	Some issues	Some issues	Yes	No	Unknown	Yes	Unknown	Free	Some issues	No
<i>Pediatric Symptom Checklist</i>	Yes	Some issues	Yes	Some issues	No	Yes	Yes	Yes	Yes	Yes	Free	Yes	Yes
<i>Strengths & Difficulties Questionnaire</i>	Yes	Some issues	Yes	Yes	Some issues	Yes	Yes	Yes	Yes	Yes	Free	Yes	Yes
<i>Child Health & Illness Profile</i>	Yes	Unknown	Some issues	Yes	Some issues	Yes	Yes	Some issues	No	Yes	No	Yes	No
<i>Behaviour Evaluation Scale</i>	Yes	Unknown	Some issues	Some issues	Yes	Yes	Yes	Yes	Yes (Short form only)	Yes	No	Unknown	No
<i>Child Health Questionnaire</i>	Some issues	Unknown	Some issues	Yes	Some issues	Yes	Yes	Yes	Yes	Yes	No	Unknown	No
<i>Clinical Assessment Package for Risks & Strengths</i>	Yes	Yes	Yes	Some issues	Some issues	No	Yes	Yes	Yes	Yes	Free	Yes	Yes
<i>Health of the Nation Outcome Scale for Children & Adolescents</i>	Some issues	Some issues	Yes	Some issues	Some issues	Yes	Yes	Some issues	Yes	Yes	Free	Some issues	No
<i>Ohio Youth Problems, Functioning, & Satisfaction Scales</i>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>Youth Outcome Questionnaire</i>	Yes	Yes	Yes	Some issues	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes

CHAPTER 7 – STUDY FOUR: CONSULTATION AND SELECTION OF AN OUTCOME MEASURE

Two important benefits of consulting practitioners about potential outcome measures are to (1) obtain feedback about the measure's practical utility and feasibility in everyday practice, and (2) include practitioners in the selection process, as the sense of frontline ownership is likely to facilitate a more successful implementation. This was particularly pertinent for the HCN project as routine outcome measurement was a new process for HCN and its practitioners, the measures being considered were not widely used in New Zealand, and they were not developed for use here.

Before seeking the views of HCN practitioners, the short-listed measures were considered by a group of clinical psychology experts who recommended which were most suited for HCN, and therefore to be presented in Part II of the practitioner focus groups. The purpose of taking the measures to the focus groups was to obtain feedback on feasibility and usefulness. The two phases of consultation for Study Four are described next, followed by the final selection of an outcome measure for HCN.

Expert consultation

Members of the HCN Outcome Measurement Project research team (Clinical Psychologists and researchers) independently considered the reviews of each potential measure before discussing as a group those that should be recommended for presentation to practitioners for feedback. A summary of the discussion follows.

Expert discussion of the short-listed measures

A drawback common to all 17 measures reviewed was the lack of data/research about their use or appropriateness for the New Zealand context. A caveat for all six short-listed measures was that practitioners would need to be assured of their suitability for use here before they were implemented in routine practice. To recap, the measures were,

- Clinical Assessment Package for Risks & Strengths (CASPARS)
- North Carolina Family Assessment Scale (NCFAS)
- Ohio Youth Problems, Functioning & Satisfaction Scales (Ohio Scales)
- Pediatric Symptom Checklist (PSC)
- Strengths & Difficulties Questionnaire (SDQ)
- Youth Outcome Questionnaire (YOQ)

The NCFAS was thought to be the best of the family measures reviewed, a domain of well-being highlighted by the HCN Advisors and practitioners as vital for tracking with HCN young people. It had the advantage of reducing subjectivity with the definitions booklet and offered a broad view of the young person's family situation. Its comprehensive coverage included the parent's functioning in the family, relationships between family members, the material environment, and the well-being of the young person. This meant NCFAS was more likely to capture important information for a wider range of young people than measures focussed specifically on one area of family functioning. Using the measure as designed – at intake and closure – could be problematic for HCN who wished for more frequent, routine administrations. In addition, because NCFAS was clinician-rated, opportunities to obtain the perspective of the young person and their parent/caregiver were missed. Despite these drawbacks, on balance, NCFAS was seen as the most suitable measure of family functioning.

The PSC was one of the briefest measures reviewed, indicating good practical utility for time-pressured practitioners. It had a large research base and was psychometrically sound. The PSC is quite mental health focussed, as indicated by the term “symptoms”

in the title, which might limit its general applicability for HCN. The multiple informant nature allowed use with a wide age range and subsequent triangulation could minimise the impact of missing data. Multiple perspectives could also be used therapeutically where differences or similarities might affect the direction of later intervention.

The Ohio Scales was reviewed very favourably by the expert group. Because it was purpose-developed as a global routine outcome measure for 5-18 year olds, it was the most widely applicable measure reviewed. The downside of this was the risk that it might lack sufficient detail for tracking a very specific problem or area of functioning if that was necessary. Unfortunately, given the highly diverse nature of HCN clients, this was true of all measures reviewed. The Ohio Scales was seen as very user-friendly, with an attractive layout and simple instructions. Like the PSC and others, it had parallel forms available for the youth, parent/caregiver, and clinician to complete. An advantage of this measure was the coverage of symptoms, functioning, satisfaction, hopefulness (a quasi quality of life scale), and living environments, in line with literature on outcome measurement (e.g., Koch, et al., 1998). It was the broadest of all measures reviewed, which added to its appeal especially given its relative brevity.

The SDQ also had multiple informant forms and spanned a large age range, although the Ohio Scales was more comprehensive in content. Like the Ohio, the SDQ is not exclusively focussed on the young person's problems, which was viewed as a strength because of the importance of being alert to both decreases in negative aspects of functioning and increases in positive aspects. The SDQ did this openly by giving scores for both strengths and difficulties, which would be useful for monitoring progress in both areas. In the focus groups, Advisors and practitioners had indicated HCN clients were likely to show change more slowly on an outcome measure than other young people, so the 3-point rating scale might lack the sensitivity to detect small changes, or not be fine-grained enough to use frequently.

The CASPARS was another taking a balanced approach to outcome measurement. The measures in this battery were explicit in this, as first, a decision is made about whether each item is a risk or strength and then the intensity is rated. As with the SDQ, two scores are obtained and could be used to track the type of change as well as where this had occurred. Like the NCFAS, CASPARS was completed by the clinician, so the view of the client and those close to them were not integrated into the overall picture. In addition, unfortunately, there was not the established body of research

behind it that supported the use of many other short-listed measures. For this reason, it was not recommended by the expert group.

Finally, the YOQ had a large amount of supporting research, was highly regarded internationally, and widely used. It was one of the longest short-listed measures, in terms of quantity of items, although they were short statements and the developers indicated administration time was around seven minutes. The YOQ contained items for a large range of issues, but some might be seen infrequently with HCN clients. Unfortunately, the YOQ was limited to use with adolescents.

Expert recommendations

All short-listed measures had positive and negative features identified in the expert group discussion. The research team felt that all measures were potential useful, however, the CASPARS lacked the research support to be recommended for HCN use at present. Aside from this, the team recommended the remaining five short-listed measures be taken to the practitioner focus groups for feedback.

The team felt that two short-listed measures could be used together to measure outcome across as many of the well-being domains as practical for HCN teams. The NCFAS could be useful where interventions were expected to improve family functioning and relationships. To supplement this with information about change in the well-being of the young person, a general measure including areas of strength and need could be used. The group thought the Ohio Scales provided that information as well as giving older youth an opportunity to contribute their perspective on their well-being and the success of the intervention plan. HCN teams could then consider multiple perspectives and integrate these to inform future work with the young person. Both measures were able to be administered by any member of the team and were not burdensome in terms of administration, scoring or interpretation time. This would allow the workload of routine outcome measurement to be shared by all and be incorporated into the everyday work of the practitioners involved.

The Ohio Scales was intentionally quite broad to give a global picture of outcome and therefore had superficial coverage of specific issues. It may lack detail if HCN teams needed to track a very specific problem, behaviour, or area of functioning, although the diversity of their clients made this issue practically impossible to avoid. Therefore, the best result was seen to be to offer a measure that was sound and could give a good

general picture of where the young person was currently, compared to a short while ago. The research team encouraged teams to supplement this with detailed measures on a case-by-case basis depending on the intervention, client's needs, and resources available.

Although the combination of NCFAS and the Ohio Scales was recommended by the team as most suitable for HCN in theory, consultation with practitioners about the best measure in practice was necessary before the final selection was made.

Practitioner consultation

Focus groups: Part II

The consultation with HCN practitioners was undertaken in the focus groups described previously (Study Two). The second part of each focus group involved presenting the short-listed measures for feedback about suitability and feasibility in practice. Participants were sent sample copies of the measures in advance (see Appendix C), so they could read and consider each before arriving. Based on the discussion of the research team, five measures were presented to the participants:

- North Carolina Family Assessment Scale (NCFAS)
- Ohio Youth Problems, Functioning & Satisfaction Scales (Ohio Scales)
- Pediatric Symptom Checklist (PSC)
- Strengths & Difficulties Questionnaire (SDQ)
- Youth Outcome Questionnaire (YOQ)

For each measure, hard copies of all informant forms and supplemental material (where applicable) were supplied at the focus groups. Recall that there were five groups: two in Auckland, and one each in Palmerston North, Nelson, and Christchurch.

Practitioner discussion of the short-listed measures

Each focus group found that at least one of the measures presented was to their liking. However, there was no clear unanimity about the most preferred measure. On the other hand, there was universal agreement about which measures were not suitable. The feedback for each measure is summarised below, with direct quotes illustrating the key ideas expressed by participants.

North Carolina Family Assessment Scale

There was general support for this measure, but suggestions for alterations to increase its utility were frequent. One common suggestion related to the potential impact on validity of recording post-intervention ratings right next to the pre-intervention ratings. The following comment captured the view also expressed in other groups.

Christchurch

I also think it would be really valuable to have one at intake and one at closure as opposed together so that you are not influenced in any way, shape or form by what's been there before.

However, this concern was not universal, with one participant in Palmerston North noting the practical ease of having the comparison easily visible. Participants highlighted the benefits of using the measure more frequently than just pre- and post-intervention. In particular, participants recognised the clinical utility of using the measure to track change over time, and how the intervention could be enhanced by its regular use. For example,

Christchurch

Yes, because if something's gone wrong in the middle you can look at it and say, "Oops, something's gone wrong here, how can we fix it? What can we put in place? Do we need to put more support in place in a certain area...?"

Other concerns related to the validity of the measure were about relying solely on clinician ratings. Participants noted disadvantages such as reducing the usefulness of the data and preventing the therapeutic opportunity to collate multiple perspectives. This was linked to a strong desire for the direct involvement of families and young people in completing measures. The dialogue that follows illustrates the sentiment in a number of groups.

Palmerston North

X: [Reading one item] "Relationships between parents and caregivers"; they might think they're completely functional and we might think that they've got a really disjointed way in which they deal with and cope with their child.

Y: ...we had Mum coming saying, "Dad's drinking heavily, I'm frightened of him" ...and we had Dad saying, "Things are great, I'm managing, I'm fine, everything's going smoothly", so you straight away have a few difficulties with how you fill out your form.

X: And yet the idea of getting the parents to do one of these, versus us doing one, you know? That'd be an interesting exercise, to be fair.

Y: It would be good to have a form you could give to different people and collate. You'd have a better idea too about what was happening at the end, because you'd see whether the change was showing across all of them or whether it was in only one environment. I think that'd be good.

A few participants discussed the length and complexity of the measure as potential feasibility problems for busy practitioners. One suggested this might lead to less careful ratings being made, which would compromise the reliability of the measure.

Palmerston North

Yip. Either it's too time consuming, or are we going to do a hash job of it?

While some participants were cautious about the measure, because not all subscales or items were applicable or useful or its design rendered it open to bias, there were more positive comments than negative at all focus groups. From two,

Auckland 1

I can see that working. It's a broad, general sort of measure; that's okay.

Nelson

I think it could have some quite good stuff around the situation for the family and improvements in family functioning...it's whether you could use something like this and also have some individual stuff around the specific behavioural intervention...yes, this is probably quite useful.

Interestingly, the bulk of the discussion about the NCFAS across groups was not about the content of the measure. Without prompting from the facilitators, the discussion was dominated by ideas about how the measure could be improved, which implied that in its current form, the NCFAS was not an ideal solution for monitoring outcome.

The Ohio Youth Problems, Functioning, and Satisfaction Scales

Of all the measures presented to the focus groups, the Ohio Scales was the most positively received. A representative sample of comments illustrates what participants liked about this measure,

Auckland 1

I like this a lot, the wellness part at the back. It's the individual answers that is good evidence as far as I'm concerned...I'll take this one!

Auckland 2

As a test, I think it's quite concrete. I quite like it. It gives some very clear information. It's quite good.

Palmerston North

Yip, and this was the one my client I was sitting down next to today decided she wanted to fill in...I was sitting down reading them outside and she sort of had a look and asked me what I was doing...and she decided that she wanted a copy to fill out.

The combination of both problems and functioning aspects were appreciated by participants. A link to the HCN well-being domains was made by a number of participants, who felt it had good coverage of a broad range of areas. In addition, it was seen as relevant to the important issues in HCN cases, for example,

Auckland 2

Facilitator: ...how closely would those things tie in with...what you want to achieve?

X: "Getting along with friends" definitely. "Getting along with family", that third one, yep. Yes, number 4, definitely. 5; yes. 6 is part of our, definitely part of it...Yes, "Controlling emotions". "Being motivated in finishing projects", not really. "Participating in hobbies", definitely, and "Recreational activities", "Completing chores", yes. "Attending school"; yes...Everything. Everything there.

Y: They're all part of the plan in all those different areas.

There were positive comments about the simple language of the measure, but queries about some of the USA terminology. One participant stated that young people would not complete any rating scale longer than nine items, which appeared related to a wider discussion in other groups about the difficulty encouraging clients to complete measures. Nevertheless, most participants seemed to hold the contrary view, that the Ohio Scales was very well designed, engaging, and would be readily accepted by young people, as illustrated by these comments:

Palmerston North

It's worded in a way teenagers can read it and understand, more so than [the measures already discussed].

Auckland 1

...often [young people aren't] very interested in participating, but they'd certainly answer a form like that.

The satisfaction scale prompted lengthy discussion about the importance of having input from both the parents and the young person into the intervention plan. This was seen as another strength, as seeking their view on how the intervention was working was highly valued. In addition, the parallel forms were recognised as another strength of the measure. One participant reiterated the view that completing such measures could provide a useful framework for engaging with family members. In particular, that the Ohio Scales had items that could be empowering for parents who would be able to see some objective indicators of positive change and that their son/daughter had some positive attributes.

Nelson

Often in terms of where I sit, a lot of the plan will work or not work on the parents' participation...so I think having an evaluation measure that actually actively includes the parent is more meaningful.

In addition to offering multiple perspectives, the therapeutic benefits of parallel forms were discussed. Participants saw the identification of differences and similarities in ratings as potentially useful for the ongoing intervention.

Pediatric Symptom Checklist

This measure was generally given little real consideration, as evidenced by the short amount of time participants spent offering comments about it in every focus group. The general opinion was that the PSC was too limited in its focus and therefore of little use. It was suggested as the type of measure so easily placed within the health domain that although it might be completed by health agencies as necessary, would not be used routinely across HCN clients. Close inspection of the transcripts showed no positive comments were made about the PSC. The following dialogue is illustrative of the discussion in other groups also.

Auckland 2

X: I'm not too keen on this. It's a bit too deficit based. If you gave this to a youth it would be just like saying, "How terrible are you? How many of these terrible things do you have?" "Oh, yeah, all of them. I'm terrible!"

Y: And a parent filling this in as well...whatever their perception of their child is, they might go, [acts out an exaggerated sigh], and go off and off and off.

X: Yes, and you can actually worsen the whole situation just by them completing this. No, I don't like [the measure].

Z: Yes, I would agree with that.

Participants also expressed confusion about the interpretation of the items, which has implications for how accurate the ratings would be. There was discussion about what several items referred to and suspicion that they may have a hidden meaning. The way the items were worded was also criticised. For example, the phrase, “too much” was deemed too ill-defined to answer accurately. Another criticism in relation to the wording for the youth self-report was that adolescents would resent the term “child” being used to describe them, as explained by these participants,

Palmerston North

X: ...you wouldn't get me asking kids that age whether they fight with other "children". You know, they don't see themselves as a child...and they wouldn't like being asked that.

Y: And "Acting younger than children your age", that's going to get a big— If you're 17 you don't really think of yourself as being a child.

Participants expressed concern that the PSC appeared more a DSM-IV-style diagnostic tool and therefore was too limited in scope and more suited for initial assessment than monitoring change over time.

Strengths and Difficulties Questionnaire

The discussion of the SDQ was interesting as it received a high level of approval from some, and little from others. Some individuals were enthusiastic supporters and current users, although some current users felt it was useful for assessment but not for outcome monitoring. Approval was for the general format of the scale items, its brevity and multiple informant design, and the relevance of its content. Participants also liked the inclusion of a number of positive/strength items that moved the measure away from a simple focus on the identification of deficits. This quote sums up the sentiment of proponents of the SDQ,

Auckland 2

[Young people] like filling them in...and I think it is user-friendly and quite unobtrusive in a way, if that's the right word...I think for parents it's quite good because when you say tell me about what the difficulties are, they often forget a lot of the things they do [well].

However, other participants felt the breadth of problems covered was limited and in all groups, concern was raised about the reading ability required for the youth self-report

being too high. A number of participants across groups felt the measure was not detailed enough to meet their needs, for example,

Nelson

It's so general it wouldn't tell you those sort of specific things...[The content] is not the nitty gritty of what the HCN funding is for and what we're hoping to make a change in.

In particular, there were doubts about the suitability of the SDQ for measuring outcome with HCN given the nature of the items and the limited range of the 3-point response scale for tracking change with complex cases.

Nelson

A lot of them would stay as "Somewhat true" because some of the behaviours you're dealing with...yeah, it doesn't give a big scale...

Youth Outcome Questionnaire

Despite its widespread use internationally and strong research backing the YOQ did not receive unequivocal support from participants in these focus groups. A common reaction was that it was potentially acceptable, but not when compared to other measures. This comment was present in most groups in various forms,

Auckland 2

Its okay, it could serve a purpose. It wouldn't be my first choice.

On the other hand, some participants appreciated the greater diversity of content in the YOQ, particularly regarding safety issues and substance use. There were also comments regarding the developmental suitability of the measure and its "mental health" feel. Some talked of the impact this would have for youth completing the form, for example,

Christchurch

...and it's like, [young people] read everything in, "I have headaches", "I feel dizzy" – "What do you want to know that for?...do you think I'm mental?" It always comes back to that; abnormal and mental is the worst thing out.

Despite concerns about the wording and suitability of some specific items, it was viewed as possibly being a useful part of a more general assessment protocol. Aspects of the measure were viewed positively, such as having a greater number of items, which was predicted to ensure a higher chance of overlap with the intervention goals. Also, the 5-point response scale was thought to increase its sensitivity to

change. However, these views were not present in the Palmerston North discussion, particularly regarding the “hundred million questions that ask them if they’re bad”. They were concerned about the length and the chance of items being missed as a result, the problem-oriented wording, and felt it did not cover areas relevant to HCN adequately. With regard to the response scale appreciated by other groups, one participant commented,

Palmerston North

I don't like the scoring part of it. I think it's too complex for kids having the five possible responses and words like "Rarely" and "Frequently". Most of the kids I work with wouldn't like that sort of language...

In contrast, Auckland participants were more favourable disposed to the measure, although not completely convinced of its suitability, as illustrated here,

Auckland 1

This is usable. This is really usable. But whether it is targeting the things that are specified in the [intervention] plan is another matter altogether...this is not inappropriate, it's just not highly appropriate.

More than any other measure presented, the YOQ triggered discussion in every focus group about the difficulty they face obtaining candid answers from young people and the effect on the reliability of the data. This issue appeared more salient in the discussion of the YOQ due to the sensitive and personal nature of some items. For some the difficulty was inherent in the measure or its content, for example,

Auckland 2

I can't imagine saying, "I have trouble with wetting or messing my pants" and my kid going, "Oh, always!"

Palmerston North

Mm, because some of these questions, kids don't know what other kids their age are like. So if you say..."I have more fears than other kids my age", how do they know that? They don't have a clue; some of them have no contact with other kids.

For others the difficulty was the disposition of the young person or a product of rapport with the practitioner, for example,

Palmerston North

The danger with [self-report forms] though, is the children, my two young people that I can think of, one likes to make things up...She's got

a different focus every week, so if you did it each week with her you'd get completely different answers.

Christchurch

...I'm not sure they're very good at self-reporting. I'm not sure that the ones I've met are going to say much about their drug use or whether they're disorganised. They might, to someone that they really know very, very well and know that they know about them, but that would have to be a pretty close person to get a genuine response.

In addition, some of the same items were thought to be difficult for a parent to answer, either because they may not know or want to know. Most often, these items were related to substance use, suicidality, or sexual behaviour. A solution was proposed in Christchurch that focussed on the enhancement of practitioner skill in administering measures. Who and how the measure was presented were considered crucial for engaging the client and receiving accurate information and, therefore, useful data.

Practitioner recommendations

All focus groups found something about at least one of the measures presented that they liked. It would be fair to summarise their responses as cautions rather than negations. In short, the most widely preferred measure appeared to be the Ohio Scales, followed by the SDQ. There was no support for use of the PSC, and the NCFAS and YOQ had mixed reviews.

On the whole, participants liked measures whose items were relevant to the presenting issues of their clients and goals of their interventions. In addition, having a balance between items for problems and strengths, and items for symptoms and functioning was considered important. Most participants expressed appreciation for measures with forms for multiple informants. There were mixed views about the value of asking youth to self-report. Although most thought this was vital for both obtaining information and enhancing engagement, some felt it was unlikely young people would participate or could be honest. Participants valued measures suited to repeated administrations over time, and those that were brief and simple in design.

Despite the positive feedback about some measures, there was universal concern about the language use in all measures. This suggested that adoption of any outcome measure requires careful editing to ensure the language was appropriate for New

Zealand. Another theme in the feedback was that while the measures were generally suitable for most clients, they were specifically suitable for fewer clients, and the data collected needed to be considered with this in mind.

Summary of expert & practitioner consultation

The expert and practitioner feedback about the short-listed measures resulted in similar conclusions. Both groups viewed the Ohio Scales favourably, citing very few drawbacks to this measure. The structure of the measure and the content of the items were seen as the main advantages, which were linked to its feasibility and usefulness for routine practice. It was the only short-listed measure to include satisfaction with services questions: a real strength that was well received. The language and layout was perceived as simple and user-friendly and was another point in its favour. The SDQ was also regarded well overall, but both groups took issue with a lack of content breadth and the restricted room for change on its 3-point rating scale. These issues were felt to limit its applicability for monitoring change with HCN clients, even though its ease of use and brevity were highly praised.

Although the expert group noted the research strength of both the YOQ and PSC, these measures were not preferred by the practitioners. They were both seen as deficit-oriented and somewhat clinical, and practitioners worried this would be detrimental to rapport with clients. This was interesting given the widespread use of these two measures in research and practice around the world. Perhaps these measures suit a mental health setting, but do not fit well in a multidisciplinary team where the interests of a wide range of professions are represented.

The conditional support of the NCFAS by the research team was echoed in the focus groups with practitioners. As the only measure of family functioning to be short-listed, it was really the only option if changes in this domain were important to track. Both groups liked the strength/risk format of the rating scale, the coverage of important issues, and the emphasis on relationships as well as practical aspects of family functioning. Conversely, both groups were concerned about relying solely on practitioner ratings and the complexity involved in completing the measure.

Even with such high levels of agreement between expert and practitioner groups, these results should be considered in context. A limitation to the practitioner feedback stems

from the low participation rate. No information was available about whether the participants in these focus groups were an adequate representation of HCN practitioners in general. Certain professional disciplines or client groups may have been over- or under-represented, which could have altered the nature of the feedback given. In addition, the dynamics and measurement experiences of the HCN teams represented by participants could have influenced their views. For example, practitioners from teams who have successfully used measures might have more interest in taking part in this type of project, and with that knowledge base to draw upon, could be inclined to concur with the expert group more so than practitioners without experience of outcome measures.

Support for the validity of these results, in spite of the sample size, came from the correspondence with the literature. Bickman and colleagues found that (1998) clinicians valued user-friendliness, relevance, and breadth of content in an outcome measure and these attributes have been recommended by experts as ideal (Lambert, 2010). The present results add to this literature by obtaining feedback on the acceptability to clinicians of specific measures in relation to one another. This allowed greater understanding of what common features are important to practitioners and gave some insight into how they weigh these features in measure selection decisions.

Experts and practitioners expressed opinions that were highly congruent despite approaching the measures from quite different points of view. Although only a small proportion of HCN practitioners took part, their opinions were consistent with prior research on the useful features of outcome measures. Of the five outcome measures that were considered by both groups, the Ohio Scales emerged as the most preferred for routine use by HCN teams.

Selection of an appropriate outcome measure

HCN requested this investigation to find suitable outcome measures for use with their young people and the four studies described so far were undertaken for that purpose. To recap, this included an extensive search to identify and short-list potential measures, a postal survey and nationwide focus groups with HCN practitioners about their views on routine outcome measurement, expert consideration of the short-listed measures and feedback from practitioners about these. From these studies arose a number of issues to consider for the introduction of outcome measurement in

multidisciplinary services, such as HCN. These were important to consider alongside the final recommendations for the selected measures.

Issues to consider for implementing routine measurement

The pivotal influence of the buy-in of practitioners for successful implementation was highlighted by remarks made in the survey and focus groups. An element of caution and unease about introducing measures was present, which must be settled in the early stages of implementation. Lambert and colleagues (2005) explained that outcome measurement introduces a degree of transparency to what is usually private work between the clinician and their client. This may be one reason for practitioner discomfort with the notion given the focus group comments about practitioners feeling the use of measures was an intrusion on their professional work and autonomy.

This resistance to increased transparency was also apparent when participants drew attention to their ability to judge progress and their clients' needs without the use of formal measures. Clinicians' overconfidence in their ability to detect poor client progress or outcome also acts as a barrier to the introduction of routine outcome measurement (Lambert, et al., 2003). There is strong evidence that this confidence is not warranted (for example, Hatfield, et al., 2009; Meehl, 1954; Sines, 1970) and that outcome measures can be a valuable adjunct for improving clinical decision-making (Hatfield & Ogles, 2006; Lambert, 2010). Therefore, if the best interests of clients are to be served, practitioners would benefit from information about how routine outcome measurement can enhance their work with clients.

Concerns of clinicians identified in previous research seem to apply to New Zealand practitioners too. These included additional administrative burden, a sense that outcome measurement is unnecessary and not clinically helpful, inadequate knowledge, training and resources, concerns about the reliance on a single measure for decision-making, and potential interference with practitioner autonomy (Bickman, et al., 1998; Garland, et al., 2003; Hatfield & Ogles, 2004; Hodgetts, et al., 2005; Merry, et al., 2004; Rock, et al., 2001). Of these reasons not to use measures, practical issues have been found to be rated by clinicians as most important (Hatfield & Ogles, 2007), so the choice of an effective and efficient measure and careful preparation for its introduction to practice are vital.

Final selection and recommendation

The extensive search for a suitable outcome measure found none was ideal or ready for immediate introduction. This was not surprising since previous similar reviews have arrived at the same conclusion (in the UK for example, Hunter, et al., 1996). No outcome measure was found that was developed in New Zealand and appropriate to the broad range of HCN young people, so further work to adapt and validate any recommended measure was inevitable. This notwithstanding, two measures were identified as meeting the needs of HCN, the multidisciplinary nature of the teams and the diversity of clients most closely. These were the Ohio Scales and the NCFAS. Together, these measures offered a breadth of content, psychometric strength, clinically useful information, acceptability to practitioners, and practical utility that was not matched by any other combination of short-listed measures. However, both were developed for specific services in the USA and while this means there is evidence of use in real world practice, no evidence could be found for their appropriateness for use with New Zealand young people. They were recommended in the final report to HCN with this caveat.

Exploring the appropriateness of both measures in a comprehensive manner was not feasible within the scope of the HCN Outcome Measurement Project. Because the Ohio Scales was potentially applicable to a large variety of young people and had the potential to be of use with multidisciplinary and psychological services outside HCN, it was chosen as the focus for the remainder of this research. In addition, this measure was identified as the best option by Bickman and colleagues (1998) who were contracted to recommend an outcome measure for routine use in Australian CAMHS. At that time however, the Ohio Scales was not well known and not in use in any large-scale manner so not implemented in Australia. Since then the situation has changed and the Ohio Scales is in widespread use and more research has provided further support. Consequently, the present research took the opportunity to explore the potential use, identified by Bickman et al., of the Ohio Scales in this part of the world, which could have implications for routine outcome measurement in CAMHS in New Zealand and – given our many societal similarities – perhaps across Australia too.

CHAPTER 8 – ADAPTING OVERSEAS MEASURES FOR LOCAL USE

The next two chapters describe the last two studies of this research, which looked into the need to adapt the Ohio Scales for New Zealand use and collect local normative data. While time and resources did not permit full investigation, the intention of this preliminary work was to provide enough information to make evidence-based recommendations on what further work was required. This chapter prefaces these studies by providing a full description of the Ohio Scales and background on the process and issues involved in assessing the appropriateness of measures developed for use elsewhere.

The Ohio Youth Problems, Functioning, & Satisfaction Scales

Development and current use

The Ohio Scales (Ogles, et al., 1999) was developed specifically with use for routine mental health practice in mind. The developers' intention was to create a measure that adequately balanced practical utility and psychometric rigour (Ogles, et al., 1999). Examination of the outcome measurement literature was the basis for the conceptualisation of the five characteristics the Ohio Scales embody (Ogles, et al.,

2000). To summarise, these were to (1) be pragmatic in terms of time, cost, and utility, (2) require minimal training to interpret and give useful results for families, (3) include input from multiple sources, including parents and young people, (4) cover multiple content areas, and (5) be psychometrically sound. To achieve this, the developers preferred a single robust measure addressing several constructs and using parallel forms rather than a battery of separate measures.

Items were generated after consulting the criteria for relevant disorders in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV; American Psychiatric Association, 1994), a list of the most common presenting issues in their region, the results of their survey of key stakeholders, and a number of well-known mental health instruments. The draft items were reviewed by service and clinical managers and parents before being finalised. The Ohio Scales in current use (and used in this research) was actually a short form of the first measure developed, but with comparable psychometric properties and increased practical efficiency (Ogles, et al., 2000), it has become the standard version.

The Ohio Scales was mandated for routine use in all public CAMHS across the state of Ohio, USA, and a substantial database of clinical data has accumulated as a result of this widespread administration (see <http://mentalhealth.ohio.gov>). Over time, this data has been used to provide benchmarks for reliable change and clinical significance, which services for young people and clinicians may find useful (Tam & Healy, 2007).

The Ohio Scales is available for a modest one-off licence fee and then can be printed or photocopied as often as required. A licence was obtained from the developer to use the measure for the purposes of this research.

Content and structure

The Ohio Scales consists of three parallel forms to be completed by the clinician (Worker form), the parent or caregiver (Parent form), and the young person (Youth form). It was designed for use with 5-18 year olds, with the youth self-report for those over 12 years, and administration time is estimated at 5-10 minutes. Each form is one double-sided A4 page with items arranged for a number of scales: Problem Severity (all forms), Hopefulness (Parent and Youth forms), Satisfaction (Parent and Youth forms), and Functioning (all forms). The Restrictiveness of Living Environments Scale (ROLES; Worker form only) is an independent instrument (Hawkins, et al., 1992) added

to the Ohio Scales for clinicians to track the living situation of the young person over the previous three months.

The Problem Severity scale consists of 20 items representing common problems with which young people present. Each is rated on a 0-5 scale for frequency in the last 30 days. Items cover behavioural (e.g., “Yelling, swearing or screaming at others”) and emotional issues (e.g., “Feeling anxious or fearful”), substance use, self-harm, and suicidal ideation. Item ratings are summed to obtain a total Problem Severity scale score that ranges from 0-100 (higher scores indicate more severe problems).

The Hopefulness scale consists of four items that cover well-being and optimism regarding parenting or the future, for the Parent and Youth forms, respectively. The Satisfaction scale also consists of four items and these include satisfaction with mental health services and involvement in treatment planning. Both scales are rated from 1-6, and are summed separately to obtain total Satisfaction and Hopefulness scale scores that range from 0-24 (higher scores indicate less satisfied/hopeful).

The Functioning scale consists of 20 items and ratings from “0 Extreme troubles” to “4 Doing very well” are made based on the impact of the problems on the young person's current level of functioning. The items cover a wide range of everyday functioning. This includes relationships with others (e.g., “Getting along with friends”), grooming, health behaviours, recreation and daily activities (e.g., “Completing household chores”), school attendance, using money, autonomy (e.g., “Doing things without supervision or restrictions”) and general psychological functioning (e.g., “Thinking clearly and making good decisions”, “Feeling good about self”). Items are summed to produce a total Functioning scale score that ranges from 0-80 (higher scores indicate better functioning).

As each scale is independent of the others, the developers indicated they can be used separately by clinicians, as appropriate for their service or individual clients (Ogles, et al., 2000). In addition, validation of the self-report form for younger children has been undertaken and provided support for the use of the Youth form for children as young as nine years old (Dowell & Ogles, 2008).

Psychometric properties

The original studies to evaluate the psychometric properties of the measure were encouraging (Ogles, et al., 2000). A later larger study supported these findings and both sets of reliability data are summarised in Table 10. Overall, excellent internal consistency (Cronbach's alpha) was found, with the exception of the Satisfaction scale in a clinical sample, which still had good reliability. Data for the community sample was not reported specifically, but noted to be comparable to the clinical sample (Ogles, et al., 2001). Test-retest reliability was adequate given the measure is required to be sensitive to change.

Table 10 Reliability Estimates Reported for the Ohio Scales' Parallel Informant Forms

Scale	Internal consistency			Test-retest reliability		
	Parent	Youth	Worker	Parent	Youth	Worker
Problem Severity						
Community sample	.89 - .90 ^a	-	NA	-	-	NA
Clinical sample	.93 ^a - .95 ^b	.93 ^b	.86 ^a - .92 ^b	.88 ^b	.72 ^b	-
Functioning						
Community sample	.93 ^a	-	NA	-	-	NA
Clinical sample	.92 ^b	.91 ^b	.91 ^a - .94 ^b	.77 ^b	.43 - .79 ^b	-
Hopefulness						
Community sample	-	-	NA	-	-	NA
Clinical sample	.87 ^b	.75 ^b	NA	.79 ^b	.74 ^b	-
Satisfaction						
Community sample	NA	NA	NA	NA	NA	NA
Clinical sample	.72 ^b	.82 ^b	NA	.67 ^b	.67 ^b	-

^a Taken from Technical Manual (Ogles, et al., 2000). ^b Taken from Ogles, Melendez, Davis and Lunnen (2001).
- Not reported. NA Not applicable.

Validity was examined (Ogles, et al., 2001) by comparing concurrent Ohio Scales scores with scores on measures of similar constructs (including the *Child and Adolescent Functional Assessment Scales*, CBCL, YSR, and *Client Satisfaction Questionnaire-8*). For the Worker and Parent forms, there were significant moderate to strong correlations for the both Problem Severity and Functioning scales. No measure was used to validate the Youth form Functioning scale, however a significant correlation was found between the Problem Severity scale and the YSR. The difference in mean ratings between the community and clinical samples were significant for parents and youth, which provided support for the discriminant validity of

the Ohio Scales (Ogles, et al., 2001). Convergent validity has also been supported for all three informant forms as scores on the Ohio Scales are associated with DSM-IV diagnostic categories as expected (Turchik, Karpenko, & Ogles, 2007).

The Ohio Scales was found to be sensitive to change in a study tracking change over a three month period (Ogles, et al., 2001). The change scores were significantly correlated with change detected on an established measure. Because of the positive psychometric evidence, the developers were confident the measure was ready for use in applied settings.

Preparation of measures for use in New Zealand

The Ohio Scales were developed in the USA and while it has been in routine use there and elsewhere, there is no published research on its appropriateness in our local context. In line with the guidelines of the International Test Commission on translating and adapting tests (see Hambleton, 2001), caution should always be exercised when considering outcome measures designed for a non-New Zealand population. Ideally, measures should be assessed for suitability and (if necessary) adapted to ensure the information they provide about clients is valid and reliable. Most of the literature regarding the cross-cultural use of measures concerns issues of language differences and translation. However, measures can require adaptation even when the language is unchanged (e.g., the country of development speaks the same language as the country of adaptation), due to the cultural, linguistic, and contextual differences underlying the interpretation of items (van Widenfelt, Treffers, de Beurs, Siebelink, & Koudijs, 2005). Therefore, it was necessary to assess the Ohio Scales for suitability before routine use could begin here, despite the fact it is already written in English.

Assessment of suitability

In order to assess the Ohio Scales appropriateness for use with New Zealand young people, the structure, items, and language were examined. This was founded in the theoretical position of “cultural universalism”: that is, the assumption that “basic psychological processes are universal for all people”, but that “culture will also influence the development and expression of psychological characteristics” (de Klerk, 2008). Pragmatically, this suits this situation as the language and many aspects of society are similar between New Zealand and the USA. Therefore, there is likely to be a degree of universal overlap in psychological processes. However, there could also

be differences in their expression based on culture, which was why the assessment of suitability was deemed useful.

Preliminary local data

Literature on adapting measures from one cultural context to another suggests a large scale field-test and re-standardisation should take place to ensure the measure's psychometric properties have been preserved (Geisinger, 1994). In addition, it is not considered good practice to use the normative data of a measure for one population if that data was collected with a substantially different population. Generally speaking, New Zealand data for child and adolescent outcome measures is non-existent, so using measures (and their normative data) developed in other western countries occurs here frequently.

To prepare the Ohio Scales for use in New Zealand, two possibilities should be considered. The first would be a field-test of the measure and comparison of the results with the original normative data. The second would be to collect complete New Zealand normative data for the measure. In fact, ideally both projects would be undertaken, with the second based on the results of the first. The first project would give information about the preservation of the original psychometric properties. However, if the measure performed differently with a New Zealand sample, then use of the original norms would not be recommended and the second project would provide a range of agencies and clinicians with robust local data to support the use of the measure. This second project would not be necessary if, in the first, the measure performed comparably with New Zealand young people as for those in Ohio.

Introduction to the Ohio Scales' studies

The two final studies of this research are outlined in the following chapters. As with the HCN studies (Studies One–Four), they have different aims, methodology, and results, so are presented separately. *Chapter 9* describes the preliminary data gathered with the Ohio Scales for a community sample of New Zealand young people in Study Five. This was consistent with the first project (field-test option) described above. The intention was to examine the psychometric properties of the measure and make recommendations about whether New Zealand normative data was required. *Chapter 10* is concerned with feedback from parents, workers, and youth on each of their respective informant forms of the Ohio Scales, which was Study Six. This feedback

was specifically regarding the cultural appropriateness of the language, structure and concepts of the measure for local use. Together, the results of these two studies were used to inform final recommendations about how best to prepare and introduce the Ohio Scales for routine practice in New Zealand.

CHAPTER 9 – STUDY FIVE: PRELIMINARY LOCAL DATA FOR THE OHIO SCALES

Measures need to have the appropriateness of their use checked in the context they are to be used in, if that context differs from the one in which they were developed (Geisinger, 1994). This can be done by field-testing the measure with a sample recruited from the new context. First, it gives a gross indication that the measure is measuring the constructs of interest in the same way. Second, any major differences in the scores of the two populations might represent important language, cultural, or contextual differences. Third, major differences in scores might represent real differences in the two populations (e.g., in the base rates of a particular disorder/problem).

Many of the psychometric measures used in New Zealand services for children and youth were developed in other western countries and are used with the assumption that they are appropriate for young people here, although this may not be the case (e.g., Black, Pulford, Christie, & Wheeler, 2010; Godfrey, et al., 1986). At the time of reviewing the shortlisted measures for the HCN project, no evidence was found that they had been normed for use here or had data gathered to provide a validity check. Study Five was intended as a step towards addressing this issue by checking the appropriateness of using the Ohio Scales in New Zealand.

In order to do this, data from a sample of New Zealand young people was compared with existing data from a similar sample in the USA. Extensive USA clinical data are available for the Ohio Scales; however, given the possibly confounding differences between our two mental health systems, comparing community samples was preferred for the purposes of this study. There were two sets of data from Ohio available for comparison with the present results; (a) the original data published in the technical manual (Ogles, et al., 2000), and (b) data from a later more extensive investigation (Ogles, et al., 2001). The original psychometric investigation of the scales used two community samples. However, for one sample only the Problem Severity scale was administered (N = 43) and the other had data for the both the Problem Severity and Functioning scales, but only for the Parent form (N = 33). The later study collected data from three groups: parents of high school students (N = 285), youth (N = 297), and parents of primary school students (N = 225), using both scales. This larger and more complete data set was selected in the present study for comparison with the New Zealand sample.

Benjamin Ogles, developer of the Ohio Scales, reported the community sample for this comparison data was recruited through two schools in a low socio-economic area of a small provincial city in Ohio (personal communication, June 2009). The purpose of his study was to confirm the validity and reliability of the Ohio Scales. To minimise potential confounds, the New Zealand sample was drawn from schools of a similar socio-economic level in a provincial city in New Zealand. Given the widespread use of USA-developed measures in New Zealand and the similarity between our two countries, the New Zealand sample was not expected to score significantly differently to the Ohio sample, and the internal consistency of the measures was expected to continue to be adequate.

Method

Participants

Participants were recruited from mid to low decile schools in the Manawatu. In total, five schools distributed packs with invitations to participate, one high school (Years 9-13) and four full primary schools (Years 1-8). The exact number of packs distributed is unknown, but schools were supplied with enough to cover their current rolls. In total, 300 packs for parents/caregivers only were supplied to primary schools, with 44

completed and returned (14.6%). For the high school, 179 packs were supplied for parents/caregivers and their 12-18 year old child. Twenty parents/caregivers (11.2%) and 21 young people participated (11.7%).

There was no missing data about who made the ratings for primary school participants: 84.1% of raters were the mother and 15.9% were the father of the child. For the high school participants, 14.3% did not indicate their relationship to the young person, 52.4% were the mother, 4.8% were the father, and 28.6% identified as “Other” (all either a caregiver or other relative). None identified as stepparents.

Materials

The packs that were distributed consisted of a brief summary of the research (printed on coloured A5 paper), the Information Sheet, the Ohio Scales, and return envelopes (see Appendix D). The packs differed slightly for primary and high schools, as the adolescents had an opportunity to complete a self-report form in addition to the parent/caregiver completed form. Thus, there was an extra Ohio Scales form and return envelope, but the information supplied was essentially the same.

Permission was obtained from the developer to adapt the Ohio Scales for the purposes of this research (B. Ogles, personal communication, April 2007). Changes made included removing the Ohio Department of Mental Health logo, altering the client details section, removing the Hopefulness/Satisfaction scales, and adjusting the copyright notice. The client details section was changed to preserve participant confidentiality and be more appropriate for research rather than clinical practice. The Problem Severity and Functioning scales were unchanged from the original measure.

Procedure

Schools agreed to distribute one pack for each child at their school to take home to their parent/caregiver. Those who chose to take part completed the Ohio Scales, sealed the form in the envelope supplied and returned it to school by the end of the week. Young people aged between 12 and 18 years could complete the self-report form and seal it in an envelope separate from their parent’s form. The forms were numbered before being distributed. Parallel Parent and Youth forms were assigned the same number to allow the matching pairs to be identified in later analysis.

The scales were scored in accordance with the procedures outlined in the user's manual (Ogles, et al., 1999), and procedures for missing items (Ogles, 2009) were followed for a very small number of participants. All forms had sufficient items completed to be considered valid. The demographic details, item scores, and scale scores of all participants were entered into SPSS 17.0 for analysis.

Results

Missing data

One Parent and one Youth form had only the second (back) page completed, i.e., no demographic information or Problem Severity scale data. This may have been inadvertent, but it meant these participants' data were excluded from subsequent analysis for the Problem Severity scale. In addition, one Youth and one Parent form had only the Problem Severity scale completed.

With regard to all other participants, at least some demographic information was supplied by everyone. In a small number of cases, missing details on the Parent form were found on the matching Youth form. For one young person no parallel Parent form was received.

Of those who completed the Problem Severity scale, only one participant missed an item. Of those who completed the Functioning scale, two participants missed an item. No scales were deemed invalid so the total scale scores for each participant were calculated for analysis.

Scale performance

Reliability

Cronbach's alpha was calculated to determine the internal consistency of the Problem Severity and Functioning scales with this sample. The Parent and Youth forms were analysed separately and for both forms and both scales excellent reliability was demonstrated (see Table 11).

Table 11 Internal Consistency Estimates for the Ohio Scales with a NZ Sample

	Problem Severity scale		Functioning scale	
	<i>n</i>	Cronbach's alpha	<i>n</i>	Cronbach's alpha
Parent Form	63	.88	63	.84
Youth Form	18	.90	21	.92

The reliability previously reported (Ogles, et al., 2001) for the Parent form was .95 for Problem Severity and .93 for Functioning. For the Youth form, similar excellent internal consistency was found, .93 for Problem Severity and .91 for Functioning.

Validity

Only divergent validity was investigated statistically for the New Zealand sample. As more severe problems are likely to result in worse functioning, a negative correlation between the total scale scores was expected. Significant, strong negative correlations were found for both Parent and Youth forms (see Table 12).

Table 12 Correlation Between Problem Severity and Functioning Scale Scores

	<i>N</i>	Pearson's <i>r</i>	<i>p</i> (1-tailed)
Parent Form	63	-.67	<.001
Youth Form	20	-.74	<.001

Comparative analysis

In order to facilitate a comparative analysis, the New Zealand sample was divided into the groups for which the Ohio data was reported. These three groups were,

- Group 1 – Parent Forms: Primary school
- Group 2 – Parent Forms: High school
- Group 3 – Youth Forms

A priori power analysis suggested a sample size of at least 35 for each group was required to test for no significant difference between the New Zealand and Ohio data at the $p = .05$ level. Due to the low response rate this was not achieved for Groups 2 ($N = 20$) and 3 ($N = 21$), so the results for those groups are presented with this caveat.

Demographic information

For the total New Zealand sample, the young people rated on the Ohio Scales were 50% male and 50% female (one not specified), and represented all ages ($M = 10.38$

years, $SD = 3.85$; one not reported) and Year levels (one not reported). There was no ethnicity specified for seven young people. Where an ethnicity was reported, 32.8% identified their child as Māori, 20.7% as NZ European/Pakeha, 19.0% as European, 17.2% as New Zealander/Kiwi, 6.9% as Asian, and 3.4% as Pasifika.

Ogles and colleagues (2001) reported demographic information separately for each group, so a comparison of age and gender with the New Zealand sample is shown in Table 13. Given the social dissimilarities between the two countries, a comparison of ethnicity information was not considered appropriate or useful.

Table 13 Comparison of Gender and Age for the Ohio and NZ Samples

	<i>Parent Form: Primary school</i>		<i>Parent Form: High school</i>	
	<i>Ohio N = 225</i>	<i>NZ N = 44</i>	<i>Ohio N = 301</i>	<i>NZ N = 21</i>
Male	46.2%	54.5%	39.2%	14.3%
Female	51.1%	43.2%	52.8%	81.0%
Missing data	2.7%	2.3%	8.0%	4.8%
Mean age in years (<i>sd</i>)	8.86 (2.23)	8.20 (2.28)	14.36 (1.54)	15.15 (1.57)

The gender proportions of the New Zealand sample were reversed for the younger children, with a difference of just over 8% more boys than the Ohio sample. In both samples, there were higher proportions of female high school students, although the New Zealand data was substantially unbalanced in this regard. There was no significant difference between the mean ages of the two samples for either the primary or high school groups ($t_{\text{Prim}}(268) = -1.751, p = .08$; $t_{\text{High}}(321) = 0.882, p = .38$).

Problem Severity and Functioning

The total scores for each scale, for each informant group, and each sample are presented in Table 14.

Table 14 Comparison of Mean Scale Scores for the Ohio and NZ Samples

	<i>Ohio Mean (sd)</i>	<i>NZ Mean (sd)</i>	<i>t</i>	<i>Cohen's d</i>
Problem Severity				
Parent form: Primary school	19.48 (18.06)	8.20 (6.11)	-7.44***	0.68
Parent form: High school	24.28 (31.76)	11.16 (12.50)	-3.83***	0.43
Youth form	33.93 (29.15)	22.40 (16.60)	-2.83 **	0.41
Functioning				
Parent form: Primary school	63.38 (14.63)	66.82 (9.55)	1.98 *	0.25
Parent form: High school	62.75 (14.17)	65.60 (9.22)	1.28 **	0.21
Youth form	60.44 (13.32)	58.71 (11.85)	-0.64	0.13

New Zealand mean scale score significantly different to the corresponding Ohio score, (2-tailed):

* $p < .05$ ** $p < .005$ *** $p < .0005$

No mean scale score for the New Zealand groups was in the clinical range, as defined by the cut-off scores in the user manual (Ogles, et al., 1999). An independent samples *t*-test was conducted to test the hypothesis that there was no difference between the mean scale scores for the corresponding New Zealand and Ohio groups. Significant, medium sized differences between samples were found for all groups for the Problem Severity scale. Significant, small differences were found for the Functioning scale on both Parent forms, but the difference between the samples was not significant for the Youth form ($p = .52$).

Figure 7 shows the comparison of 95% confidence intervals for the New Zealand and Ohio samples, illustrating the size of the difference in the Problem Severity scores. The New Zealand scores are lower (i.e., less severe problems) than the corresponding Ohio scores. The upper limit of the New Zealand data does not overlap with the lower limit of the Ohio data for any of the three groups.

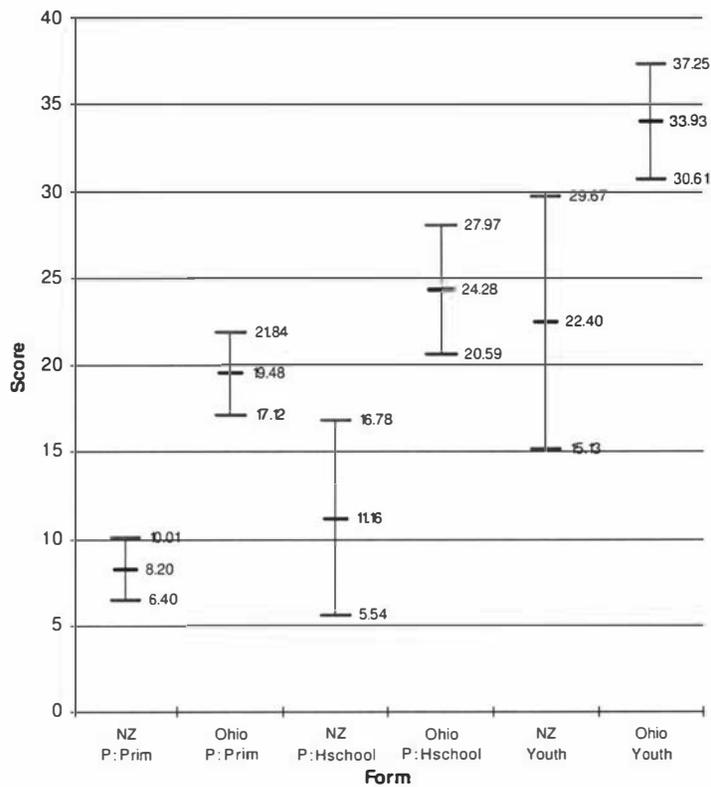


Figure 7. 95% confidence intervals for mean Problem Severity scale score for the three informant groups in the New Zealand and Ohio samples

Figure 8 shows the comparison of 95% confidence intervals for the Functioning scale scores. For these scores the comparison to the Ohio data is less clear. It appears that for the Parent forms the New Zealand scores are higher (i.e., better functioning), and lower for the Youth form. However, the intervals overlap for the primary school scores, whereas the Ohio intervals are almost completely within the New Zealand interval for both sets of high school data. The wide intervals for the New Zealand scores reflect the small sample sizes, particularly for the high school data.

No statistical analysis or comparison of scale scores by sub-groups (such as gender, Year/age, completer, or ethnicity) was conducted. This was not thought appropriate due to the small sample sizes involved. The same also applied to item level analyses.

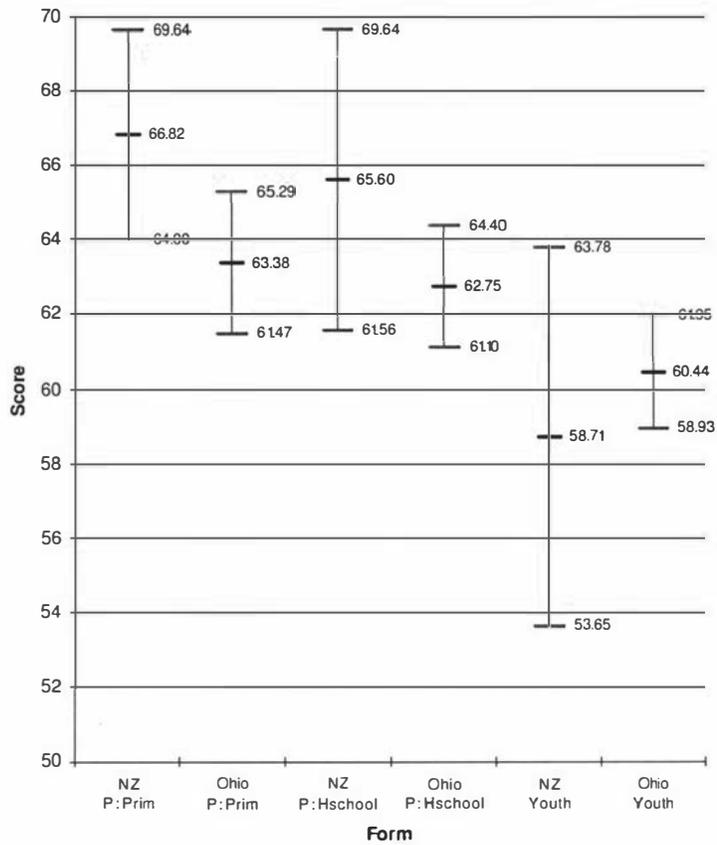


Figure 8. 95% confidence intervals for mean Functioning scale score for the three informant groups in the New Zealand and Ohio samples

Summary

In order to investigate the performance of the Ohio Scales with a New Zealand sample a comparative analysis was undertaken. A community sample of 5-18 year olds in New Zealand completed the measure (64 parents/caregivers and 21 youth), which represented a response rate of approximately 11-15% for the three informant groups (Parent form: Primary school, Parent form: High school, and Youth form). Differences between this sample and the data from a comparable Ohio sample were investigated. The internal consistency of the scales for the New Zealand groups was slightly lower than that reported for the Ohio groups, although still excellent. A strong correlation between the Problem Severity and Functioning scale scores was in the expected negative direction. Thus, there was preliminary evidence the psychometric properties of the Ohio Scales were preserved adequately, as predicted, when used with New Zealand young people. However, with regard to the predicted equivalence of the mean scale scores between the two countries, the results were considered inconclusive.

For all groups the mean Problem Severity scale scores were significantly lower in the New Zealand sample. This was also the case for parent ratings of their child's functioning, but not for youth self-ratings. A cautious interpretation is warranted as these results are limited by a number of factors and are therefore inconclusive without further data being collected. Most obvious is the small number of participants recruited, particularly in the high school groups. In those groups the sample size did not reach the level indicated for adequate power (i.e., 0.8) to detect equivalence. With a larger sample of high school students, the comparison of scores should provide more clear results. Related to this is the low response rate. Although similar methodology was used to recruit participants, the Ohio researchers had response rates of around 60% for high school and 56% for primary school aged children. Given the close proximity of the university, perhaps the New Zealand schools and families suffer "research participation fatigue", and so were less likely to take part (one school that was approached, declined to be involved for this reason). In addition, the New Zealand data were collected during Term 4 and Term 1, which several schools mentioned were hectic times for teachers and families (two schools declined to be involved for this reason). The low response rate may reflect the multiple demands on families' time. It is also possible that families dealing with high levels of distress will have chosen not to take part at such a busy time of year. If this was the case, it could account for the low level of difficulties reported by those who did participate. To build on this preliminary information, recruiting a larger sample is crucial and may be easier if done during the middle of the school year and perhaps from schools who are not often approached for research projects. This would provide a stronger foundation for clarifying these initial results.

Since the data in this study were collected, New Zealand data for the self-report SDQ have been published for a sample of 484 Auckland high school students (Black, et al., 2010). Black and colleagues found scores that differed from the previously published international normative data. In particular, they found only 4% of their sample scored in the "abnormal" range using the published cut off scores, which are set at the 90th percentile. The trend in the present study is consistent with Black et al.'s larger, more robust sample. They also noted that in comparison to the Australian normative data, male students had higher levels of peer problems and female students more emotional symptoms and higher Difficulties scores. As a result, these authors suggested practitioners be cautious when interpreting SDQ scores for New Zealand young people.

It is important to note that even if the present study's findings had argued against cross-national equivalence for the Ohio Scales, the reason for this would be unclear. If not representing a real difference in base rates in the two countries, another possibility relates to the interpretation of items. Further investigation would be useful to explore how the item language and concepts are understood by New Zealand raters as this could explain cross-national differences in the total scores. This would require looking at terminology that is less familiar here than in the USA, and whether there are differing social expectations about what is and is not considered problematic or developmentally appropriate. For the scales to be adjusted for use here, these issues should also be looked at with regard to appropriateness for Māori whānau. Although no suitable measure is ready at present, in the future a concurrent validity check with a culturally appropriate Māori outcome measure would be valuable.

It is important for clinicians to consider the issue of using measures developed overseas with New Zealand young people, particularly when using clinical cut-off scores or normative comparisons. Although the present findings were not conclusive, if further work reveals significant differences between New Zealand and USA young people's scores, this would pose a dilemma for local clinicians. There may be a risk that New Zealand young people are disadvantaged when overseas measures are used as part of the clinical decision-making process. It would be a valuable contribution to the local literature for work to be undertaken to validate good routine outcome measures in New Zealand, including clarifying the findings of this small study by recruiting a larger sample.

CHAPTER 10 – STUDY SIX: INFORMANT FEEDBACK ON THE APPROPRIATENESS OF THE OHIO SCALES FOR NEW ZEALAND

As previously mentioned, Geisinger (1994) stressed the importance of ensuring measures are appropriate and valid before using them outside the population they were developed for, even when no changes have been made to the measure. Assessing the appropriateness of the Ohio Scales prior to their administration with a New Zealand population is important and Geisinger provided a useful outline of the steps involved. Given that New Zealand is not too dissimilar to the USA where the Ohio Scales were developed, the assessment of their appropriateness was expected to be straightforward.

Geisinger suggested measures be assessed at two levels: the conceptual level and the item level. The psychological constructs assessed by the Ohio Scales are common to most western cultures, so an assessment at the conceptual level was not deemed necessary at this stage⁸. Item level assessment would enable terminology, examples, and language use to be considered appropriate for a New Zealand population, or not.

⁸ However, it is important to point out that the mental health and functioning constructs covered by the Ohio Scales might not be suitable for measuring outcome within a Māori or Pasifika service. This would require further work to adapt the measure, which was outside the scope of this project.

In addition to the actual items, instructions and the client details section should be included in the assessment of appropriateness (Geisinger, 1994).

Geisinger recommended this type of assessment involve a group of people to provide comment on the measure. These people should have suitable knowledge of the local culture/society, psychometric testing, and in this case, psychosocial practice with young people and their families. In particular, the Ohio Scales are rated by multiple informants, so each informant group was likely to provide different perspectives and feedback. Thus, obtaining feedback from parents/caregivers and young people, in addition to those with psychology training, was considered essential.

Ultimately, deciding whether the measure was appropriate for the New Zealand context relied upon the feedback received from these three groups. After considering the feedback, one of three possible decisions could be made. Van de Vijver and Hambleton (1996) described these as,

1. apply the measure,
2. adapt the measure, or
3. assemble an entirely new measure (for this study, a decision to discard the Ohio Scales).

Option One, deciding to apply the measure in its original or very similar form (i.e., New Zealand spellings and language usage), would be the most straightforward to pursue. If the assessment showed items were in danger of having a different interpretation/meaning for New Zealand respondents then Option Two, adaptation of those items, would be indicated. It was predicted that one of these first two options would be the outcome of the appropriateness assessment, and the feedback gathered could then be used in future modification of the Ohio Scales for use here.

Method

Participants

Three groups of participants were recruited, matching the three informant forms of the Ohio Scales: Workers, Parents/caregivers, and Youth. The convenience sample was recruited using a “snowballing” technique. An email was sent to 20 people inviting them to participate and/or pass on the details of the study to people they know.

Because the number of people the study details were then forwarded on to is unknown, a response rate cannot be calculated. However, 24 people took part, which suggested the study details were passed to others in a snowball effect as intended. The email briefly described the study and procedure and included three attachments (one for each informant group). Each attachment contained the Information Sheet, Consent Form, and the Feedback Questions sheet (see Appendix E). For clarity, the details for each group of participants are described here separately.

Workers

Eleven participants were in the “Worker” group, eight women and three men. To be included in this group, participants were either clinical psychologists or clinical-psychologists-in-training, in line with the suggestion of Geisinger (1994) that those giving feedback have post-graduate training in psychometric assessment and working with young people and families. If a participant met the criteria for inclusion in both the Worker and Parent groups, their level of training meant participation in the Worker group was more suitable (this applied to six participants).

Parents/caregivers

Seven participants were in the “Parent/caregiver” group, five women and two men. To be included in this group, participants needed to currently have at least one child aged between 5 and 18 years, and not meet the criteria for the Worker group.

Youth

Six participants were in the “Youth” group. There were four girls and two boys who ranged in age from 13 to 17 years ($M = 15$ years). To be included in this group, participants needed to be aged between 12 and 18 years (the age range the Youth form of the Ohio Scales is designed for), and have their parent/caregiver’s consent if they were aged 16 years or younger.

Materials

Participants received a copy of the applicable informant form of the Ohio Scales and a page of prompt questions to guide their feedback about the measure. The prompt questions were as follows (similar but simplified wording was used with the young people taking part; see Appendix E):

1. Do any changes need to be made to the questionnaire so New Zealand English is being used (i.e., spelling, grammar, and punctuation)? Examples and suggestions?
2. Should the wording be modified to reflect New Zealand's bicultural society? How?
3. Do the questions use words/descriptions that are different to what we would use in the same question here? Examples and suggestions?
4. Are there ideas/language/concepts that are very "USA", but not very "New Zealand"?
5. Anything else that might not reflect life in New Zealand as it does in the USA?

Procedure

Participants used a copy of the Ohio Scales applicable to their informant group (i.e., Worker, Parent, or Youth form), and read a page of instructions about what was required. They considered the five prompt questions about possible changes to the Ohio Scales for it to be suitable for New Zealand use. They were invited to use these questions to guide their thinking, although were not required to structure their feedback in this way. The quantity and quality of responses was thought to be maximised by offering multiple options for returning feedback. Therefore, participants were able to either write answers to the questions, give feedback orally, email a summary of their thoughts, or write/draw directly on the measure or a separate sheet. All of these options were used by at least one participant each.

Once participants had supplied their feedback, they were thanked for taking part. The individual responses were typed and collated according to informant group.

Results

The analysis of the feedback consisted of grouping similar content, which was straightforward as there was a very high degree of consistency in the responses. As the feedback was intended to provide the subjective perspectives of New Zealand informants, no statistical analysis was required or appropriate. The feedback is presented here for each section of the Ohio Scales, as many of these are parallel across informant forms.

Client details section

Most Worker and Parent participants commented on two terms in this section of the Ohio Scales: “Race” and “Grade”. The Youth agreed with changing “Grade”, however, none noted “Race” as needing to be changed. These terms were identified as requiring replacement words to suit the New Zealand context, with the most common suggestions being “Ethnicity” and “Year” at school, which were said to be usual terminology here. One participant remarked,

“Ethnicity” seems more culturally sensitive in NZ and acknowledges it’s not just about biology.

Additional suggestions from the Workers were to allow multiple ethnic affiliations, to use the biculturally appropriate “young people/rangatahi” in preference to “youth”, and to change “therapist” to “clinician”. One suggested adding “Other whānau member” to the options given on the Parent form for the relationship of the rater to the young person.

Problem Severity scale items

The most frequent feedback from participants for this scale was about the terms “curfew” and “skipping school”. Both were identified as North American idiom not used in New Zealand; rather, “agreed time to be home” and “truanting”, “bunking”, or “wagging” were suggested. Other feedback about items on this scale appeared to not be related to cross-cultural issues, but about the clarity of the items. In particular, Workers expressed uncertainty about what a “fit of anger” looks like, and challenged the notion that a young person “causes trouble for no reason”. Another mentioned that “talking or thinking about death” may not always be problematic for cultural and/or situational reasons.

“ROLES” section

The Restrictiveness of Living Environments Scale (ROLES) is only included on the Worker form, so feedback about this section was only received from one group of participants. It is also important to note that this is a stand-alone psychometric measure developed independently, and prior to, the Ohio Scales. However, if the Worker form is to be used in New Zealand with the ROLES included, consideration of this feedback would be essential.

This section received more feedback than any other did, and it was fairly consistent across participants. This participant summarised the ROLES feedback succinctly,

Mostly it is the use of words to describe specific environments...which are conceptually USA – but if these are changed, the scale would be useable.

Participants did not know what many of the living environments listed actually were, for example “Residential Job Corp/Vocational Center” or “Group Emergency Shelter”. Others were recognised, but a local term was suggested as a replacement. These included replacing “Jail” with “Prison”, “Dormitory” with “Boarding School”, “Independent living with a friend” with “Flatting”, “Group home” with “CYF home”, and “Juvenile Detention Center” with “Youth Justice Facility”. A number of participants queried the meaning of a school detention in the USA, as the wording “days in detention” implied much longer punishment than the 15-30 minutes reported as typical for New Zealand schools. To be more inclusive of Māori and Pasifika communities, new living environments were proposed as possible additions. These included “whāngai”, “living with grandparents/koro/kuia”, and “home of relative/whānau”.

Hopefulness and Satisfaction scale items

In place of the ROLES section, these two scales are included on the Parent and Youth forms, so feedback for these scales was only asked of these two groups. No participants provided feedback related to New Zealand appropriateness, but one suggestion was to use the term “hopeful” rather than “optimistic”.

Functioning scale items

A number of Worker comments about this scale were concerns about possible double-barrelled questions. Upon examination of the items mentioned, these concerns appeared based on the use of the word “and” in the item, despite the two parts of the item being conceptually related. For example, although “concentrating, paying attention, and finishing tasks” contains three parts, they all contribute to a rating of the same concept. A true double-barrelled question would tap separate concepts; such as “feeling in a low or anxious mood, or troubled by fears, obsessions or rituals” (Item 9, HoNOS-CA).

Terminology not used in New Zealand was identified by Worker and Parent participants. “Chores” and “passing grades in school” were frequently cited.

Alternatives offered were, “jobs”, “jobs around the house”, or “cleaning room and other jobs”, and “coping with classes”, “satisfactory achievement”, “making progress in class”, or “passing subjects”. One Youth participant also suggested, “Doing jobs”. There was a suggestion to change “principal” to “headmaster/mistress” and a parent wondered if the item about dating and romantic relationships was less “the norm” for New Zealand teenagers than in the USA. One Youth offered an alternative wording for this item: “...going out with boyfriends or girlfriends.”

The remaining feedback was regarding the examples of hobbies and recreational activities. Almost every participant suggested removing “baseball cards”. Alternatives such as, “art, music, collecting”, “rugby cards”, “Weet-Bix cards”, “Pokemon cards”, “reading, music, craft”, and “computer games”, were offered. Two Parent participants commented,

The whole hobby thing seemed like something from the 50s! All about collecting things!

I know this isn't what you asked but the hobbies and recreational stuff are kind of male! Where is dancing? Or gym?

It was suggested that specific sports could be listed, such as rugby or soccer. Finally an additional item was suggested, “Participating in cultural activities”, although another participant mentioned incorporating Māori examples of hobbies/recreational activities into the existing items, such as Kapa Haka.

Bicultural appropriateness

Almost all participants suggested slight changes or additions should be made to suit New Zealand's bicultural context, although one felt this was not as necessary for the Worker form as for the Parent and Youth forms. Most often, the feedback was to add the equivalent Māori term, e.g., whānau/family, tamariki/child, kura/school. A number of Worker and Parent participants suggested offering a Te Reo Māori version of the measure, and one participant thought there was room to incorporate the ideas and language from Te Whare Tapa Wha. One of the young people mentioned using “simple Māori words that all New Zealanders know, such as ‘Kia ora’ etc.”

General feedback

Some participants felt the word “Problems” was judgmental and Workers and Parents

preferred “Difficulties” and “Issues”, respectively. The use of “Ohio” in the title was questioned by one Worker, who felt it conveyed a sense of foreign-ness. One Youth thought the form was easy to understand and one predicted that students would lie on some questions. A number of Parents and Youth felt that the measure did not need changes to New Zealand English, although one queried the suitability of the reading age required for the Parent form.

Summary

Twenty-four clinicians, parents, and young people supplied feedback about the appropriateness of the language and concepts of the Ohio Scales for New Zealand. There were strong themes in the feedback about wording that should be altered (such as, “chores”, “curfew”, school “grade”, “race”), although the measure appeared to be conceptually acceptable to participants. Most suggestions were to alter a single word to its familiar New Zealand equivalent. The ROLES section of the Worker form received the most feedback and appeared to require the most substantial modification. Many of the terms referred to services/settings that either do not exist in New Zealand or exist in quite a different form. Uncertainty about the meaning of many terms in this section was frequently expressed by participants, so it seems that this scale requires substantial modification before it could be used in New Zealand.

Of Van de Vijver and Hambleton's (1996) three possible decisions, it was predicted that the feedback would indicate a decision to either apply or adapt the Ohio Scales for New Zealand use. In fact, the results of this study indicated both are necessary. For the ROLES section of the Worker form, adaptation and New Zealand validation is necessary, as the terminology of the scale was clearly not understood by participants. However, this is an independent scale, so could be removed from the Worker form altogether without affecting the Ohio Scales proper. Fortunately, the rest of the measure did not suffer such difficulties. The feedback indicated only minor rewording was required. Based on the findings of this study, the measure could be applied here in a very similar form to the original.

Although a convenience sample was used in this research, the views of the participants were highly convergent, so the sampling method may not have impacted on the findings. As Geisinger (1994) had recommended, a group familiar with the New Zealand context, psychometric testing, and psychosocial issues for young people was

recruited. However, all Workers had a background in clinical psychology and practitioners in other disciplines (e.g., social work or nursing) might have given different feedback. In particular, this limitation might apply to the ROLES section. Perhaps many of the services/settings mentioned would have been familiar to professions other than psychology. Although a number of participants were Māori, no separate analysis of their responses was made and therefore, no conclusions were drawn about the appropriateness of the measure for Māori. This would be a critical future research project, particularly if the measure was to be considered for use with predominantly Māori clients or services working from Kaupapa Māori perspectives.

In line with copyright legislation, the Ohio Scales (and the ROLES) cannot be altered or adapted without permission. Even though no major adaptation is necessary, adding an organisation logo or making minor changes to the measure (such as New Zealand spellings or adding words in Te Reo Māori) would breach copyright. The next step would be to seek permission from the copyright holder to make alterations as suggested by the results of this research.

Once permission to make changes was obtained, an initial New Zealand version could be drafted. Geisinger (1994) suggested an editorial review of the initial draft be undertaken where independent reviewers individually make comments regarding the changed items/sections, which are then shared amongst the reviewers, discussed, and a consensus reached about what needs adjusting (if anything). If the ROLES section was adapted, a similar process would be followed, however care should be taken to retain the original meanings of items, so as not to spoil the conceptual underpinnings of the measure. The measurement equivalence of the original and adapted items and their effect on the measure as a whole and its psychometric properties should also be investigated (see Robert, Lee, & Chan, 2006).

In addition to the development of a New Zealand appropriate version of the Ohio Scales, future research could look at drafting and testing a Te Reo Māori version. This would require extensive consultation, piloting, and evaluation, but would strengthen the usability of the measure in both mainstream and Kaupapa Māori child and adolescent mental health services.

As predicted, the results of this research indicated the Ohio Scales could be applied for use in New Zealand with only minor wording changes that do not compromise the meaning of items or structure of the measure. The inclusion of the ROLES section was

not supported however, and further adaptation of this scale is recommended before it is used by services here.

CHAPTER 11 – SUMMARY AND CONCLUSIONS

Summary of research findings

Three aims were proposed for this research and a series of six studies were undertaken in order to achieve them. The four HCN studies were a postal survey of practitioners, focus groups about outcome measurement, a comprehensive search to locate potential measures, and expert and practitioner consultation on the short-listed measures. The Ohio Scales was selected as best able to meet the requirements of a good routine outcome measure for multidisciplinary interventions with young people. The two Ohio Scales studies were a preliminary comparative analysis of the measure's performance with a New Zealand community sample, and an assessment of its appropriateness for New Zealand use. A summary of this research follows, organised to illustrate how the findings contributed to the achievement of the three original aims.

Aim 1

Provide an updated overview of potential routine outcome measures available for use with children and young people

Previous projects in the UK (Hunter, et al., 1996) and Australia (Bickman, et al., 1998) to identify the most useful outcome measures for routine clinical practice did not find any single measure ready for implementation. Since these projects were undertaken,

the field of outcome measurement has progressed and purposed-designed routine outcome measures have been developed and validated. The present research involved an updated search to locate potential measures for use with young people, with a particular focus on suitability for New Zealand practice.

The steps taken to identify and select suitable measures were similar to those used by Koch et al. (1998) and Bickman et al. (1998). The search used a conservative approach to screen the 1665 potential measures located. The screening process was guided by the HCN domains of outcome and criteria for an ideal measure, and resulted in 17 measures being considered. Brief reviews of these measures led to a short-list of five measures that were comprehensive in their content coverage, and one measure focussed on family well-being and functioning. All included useful items, had methods of making raw scores meaningful, provided data for use at multiple levels, required little administration time or training to use, and were available for minimal or no cost.

A drawback common to all was a lack of New Zealand validation. Only one New Zealand measure was located (Hua Oranga), but this was still under development and there were concerns about its suitability for use with young people. In addition, no measure perfectly met all the requirements or was applicable across all ages and client presentations. This could have been a product of the parameters of the search, although it is unlikely any outcome measure would ever meet the needs of every client, clinician, and service. Measures may exist that would come closer to ideal, but were not identified in this search. For example, measures whose use is restricted to psychologists or have a long administration time have the potential to be more comprehensive in content or more broadly applicable. Unfortunately, the practical requirements of outcome measurement in routine multidisciplinary practice preclude the use of these measures, so they were not considered.

Although this search for measures was undertaken with the needs of HCN in mind, the short-listed measures are suitable for other multidisciplinary services for young people. Egelston and Watkins (2008) recommended alternative outcome measures be considered to replace the mandated HoNOS-CA, so the results of this search are likely to be of interest to CAMHS, as the short-listed measures were reviewed more favourably than the present mandated measure, HoNOS-CA. In addition, those working in individual therapeutic settings are also likely to find the measures useful for monitoring client progress and informing treatment decisions. This search and review of measures contributes to the wider international outcome measurement literature by

reducing the chaos clinicians are faced with when selecting a measure. It also built on previous work by updating the older reviews of Hunter et al. (1996) and Bickman et al. (1998), and was the first systematic review of routine outcome measures for use with young people in New Zealand.

Aim 2

Select at least one that fulfils the requirements of a good routine outcome measure

As outlined previously, the selection of a routine outcome measure can be challenging due to the many factors that must be considered. The search and brief review of measures covered important conceptual, structural and psychometric factors to create a short-list. In order to select the “best” measure, information was gathered about two further factors: usefulness and feasibility. The results of the postal survey and focus groups of practitioners highlighted the elements of a measure that were valued and those that were problematic.

Participants valued the ability to track client progress with outcome measures, which is directly related to the usefulness criterion. As emphasised in our *National Mental Health Information Strategy* (Ministry of Health, 2005), clinicians must find the measure and its results useful or they are less likely to administer it. Outcome measures were contradictorily perceived as either too specific or too broad and this was linked to their utility as an accurate monitor for individual clients. This suggested practitioners valued a balance between having enough detail specific to their client, but enough breadth to be applicable to a range of clients. Survey participants rated client behaviour, reports from significant others, and symptom reduction as the most important indicators of progress and outcome. The information from measures was seen as being much less so. If a measure could capture those indicators clinicians value most, it would be reasonable to assume their perception of the utility of measures may be more favourable and increase their willingness to use them with clients. At a more concrete level, most practitioners wanted the content of measures to include functioning, mental health and safety concerns, and relationships with those close to the client. In addition, opportunities for parallel informant forms, repeated administration, and comment on satisfaction with the service were thought ideal in a measure.

With regard to feasibility, participants valued measures that were user-friendly, both for the client to complete and the practitioner to administer. This was the biggest barrier to practitioners using measures so was a crucial consideration in the selection process. There was an emphasis on measures being low cost, very brief, simply worded, and easy to score. Practitioners wanted to see relevance to New Zealand, and this included psychometric validation here and local normative data being available.

The short-listed potential measures found in the search was presented to practitioners and experts for consultation. They independently preferred the Ohio Scales for tracking general outcome and the NCFAS for tracking family functioning, and once the results of the survey and focus groups were considered, these were the final selections recommended to HCN. For monitoring outcome in a broad sense, the Ohio Scales met all the criteria for a good routine outcome measure. Most importantly, given the views of practitioners mentioned above, it had the widest coverage of content domains including symptoms, quality of life, satisfaction with services, and functioning, yet was still relatively brief and simple to use. In addition, its ability to detect change over time was enhanced by its 5-point rating scale and short rating period.

The search, review, and recommendation of a measure conducted for this research is the most comprehensive since the review of Bickman and colleagues (1998). With regard to identifying a good routine outcome measure, the final selection in both reviews was the same. Although both identified the Ohio Scales as a preferred measure of outcome, and it now has further research and clinical data behind it, it is not in large-scale use internationally. This could be due to the non-commercial nature of its development, which may mean a lack of funding behind validation work to support its use in international settings. This is disappointing considering it has twice been identified as the best routine measure available for use with young people.

Aim 3

Obtain preliminary information about the measure's appropriateness for use with a New Zealand population

To begin to extend the international support for the use of the Ohio Scales, further work was undertaken to investigate its appropriateness for New Zealand young people. Unfortunately, as with all the short-listed measures, the Ohio Scales has no New Zealand normative data and no validation research conducted here. This data was

found to be important to practitioners in the survey and focus groups, and is crucial to ensure best practice outcome measurement.

In line with the guidelines of the International Test Commission (Hambleton, 2001), the Ohio Scales was administered to a community sample of New Zealand young people and the scores compared to a similar sample from Ohio. No significant difference in the scores was expected and the internal consistency of the measure was predicted to remain strong for the New Zealand sample. The first of these predictions was not supported by the data, although the excellent reliability of the measure was preserved. To summarise, the New Zealand sample scored lower on the Problem Severity scale and parents rated their child's functioning as better. The results were inconclusive due to the small sample size, so no recommendation could be made regarding the necessity of collecting local normative data. Further data collection was recommended, including clarifying the reason for cross-national differences in scores if appropriate (for example, different base rates of emotional and behavioural problems, different interpretations of items, or wider social/cultural factors).

A further study was conducted to identify whether the language of the measure needed to be altered to suit New Zealand idiom and ensure the items would be understood adequately. The results of feedback on each of the parallel forms from the three informant groups (clinicians, parents, and young people) led to the conclusion that only minor adaptation was required. In the main, this included substitution of words or phrases using USA spelling or colloquialisms. However, the feedback on the ROLES section of the Worker form indicated substantial modification was required and for this reason, it should be removed from the measure until a New Zealand version of ROLES has been developed and tested. Considering the results of both studies, it would seem sensible to make the minor language changes suggested and undertake a second, larger, validation study with a New Zealand community sample.

Discussion

The ultimate aim of this research was to provide evidence-based recommendations on the outcome measure most suited for routine use in multidisciplinary interventions with young people in New Zealand. Based on previous similar projects, it was predicted that no measure would be ideal for immediate use here. This was the case, although one measure emerged as the strongest contender. The Ohio Scales met all the criteria

for a good routine outcome measure (i.e., strong psychometric profile, useful to those interested in its results, and feasible for routine use), but lacked validation or local normative data to support its use in New Zealand. Two small studies to assess the need for further validation work suggested that further data from a larger sample should be collected, although the measure itself required only minor wording changes.

The results of this research should be interpreted with the following caveats. First, while the range of measures considered during the search process was exhaustive, many potentially useful measures were excluded from the final review. Because this search and review was guided by the requirements of HCN and the nature of their work with young people, good outcome measures were screened out in this project. While not useful for HCN purposes, they may be useful for those working in different settings or requiring measures to monitor different outcomes. The measures considered in the HCN project were only those that fit a multidisciplinary context so any requiring advanced training to administer were excluded. Other measures excluded from consideration were those only for children aged less than five years, those with a narrow content focus (e.g., depression symptoms), and those that fit within a particular therapeutic framework (e.g., cognitive distortions for Cognitive Behaviour Therapy). For clinicians wishing to monitor outcome in these specific circumstances, the results of the search are not likely to provide adequate options. In addition, the practitioners who participated in the survey and focus groups were working for HCN, a very specific multidisciplinary context. This meant that some of the disciplines providing feedback would differ from those in other agencies. For example, CAMHS do not include teachers or police so practitioners in these services are likely to provide feedback on outcome measurement with less emphasis on education or justice issues. It would be a valuable future project for the search and review process to be repeated to identify the best outcome measures in these specific domains. This would allow the compilation of a “catalogue” of measures for clinicians to select from, as was suggested by participants in the present research. With this catalogue, clinicians could balance routine outcome measurement with the Ohio Scales with more fine-grained monitoring of progress in those areas their intervention is expected to target. It would also be useful to conduct focus groups with young people and parents/caregivers to elicit their opinions on outcome measurement and feedback on specific potential measures, as was done with HCN practitioners in this research. Given practitioners’ comments about having the buy-in of clients for successful outcome measurement, feedback from all potential informants would assist in selecting a measure that could facilitate the smoothest implementation and be maximally useful.

Second, the small sample sizes in this research mean the results must be interpreted with caution. In the survey and focus groups, the response rates are likely to be underestimated, as it was not possible to ascertain accurately the number of practitioners who received the invitation and declined to participate. In addition, no information was available to determine how representative the samples were of the professional disciplines working for HCN, which in turn made it unclear how representative the participants' feedback was of HCN practitioners in general. However, the views expressed were wide ranging, clear themes emerged, and the responses were consistent with the literature in this area. Taken together, this gives a degree of confidence to the assumption that the results are broadly indicative of the opinions of multidisciplinary practitioners in New Zealand. Much less confidence can be placed in the results of the study to collect of New Zealand data on the Ohio Scales. Although the data showed that New Zealand young people in a community sample scored significantly differently to those in Ohio, this may well be an artefact of the small sample size. Interestingly, the HoNOS-CA scores analysed by Trauer et al. (2004) from a very large CAMHS data set were not psychometrically robust and a recent study by Black et al. (2010) found cross-national differences in scores for the SDQ, which may indicate that New Zealand young people do indeed score differently to their international peers. In either case, based on the limited data for the Ohio Scales in the present research, it is not appropriate to generalise beyond this sample until further data is collected.

A larger scale study with the Ohio Scales in New Zealand would be a worthwhile future project. To be most time and resource efficient, this could be done in two stages. First, the Ohio Scales could be adapted according to the suggestions of the clinicians, parents, and young people in this research to create a New Zealand language version. This would then need to be piloted to ensure the adapted measure performed equivalently to the original. Without this pilot, any future comparison of New Zealand and Ohio data would be confounded by the language changes. Second, a community sample of comparable composition to the Ohio data should be collected in order to clarify the results of this study. To overcome the limitations of the present research, a sample of similar size to the Ohio comparison group should be recruited.

The results of this new comparison study would either, support the use of the measure here as is, or indicate the need for New Zealand normative data to be collected. If the latter were necessary, two types would be useful: community and clinical normative

data. Community norms would allow clinicians to estimate the degree to which their client differs from the general population of New Zealand young people. It would also allow a clinician to evaluate whether there has been any improvement after an intervention by showing that their client's scores have moved towards or into the "normal" range. It is important to note that community normative data could not be obtained from professionals providing intervention services, as most young people are not involved with such a person. This would mean that community norms could be developed for the Parent and Youth Forms, but not for the Worker Form. Clinical norms could be collected for all three forms. The aim would be for these norms to be representative of young people involved in intervention services (e.g., HCN or CAMHS). Ideally, these would consist of data from clients with a range of presenting issues. This would allow for the comparison of a young person's scores with those of the "average" client of clinical services. For both types of normative data, the sample needs to be large enough that characteristics where differences are likely to be present, such as age and gender, can be analysed. If differences are identified the sample could be split into appropriate smaller sub-groups (e.g., boys aged 5-10 years) to give more developmentally appropriate comparisons.

The main contribution of the present research to the field of outcome measurement is to provide guidance to services, clinicians, and clients in New Zealand when selecting a routine outcome measure. This research represents the first systematic and comprehensive search and review of outcome measures for use with young people in multidisciplinary settings in this country and an update of previous international reviews. The results are directly applicable to clinical practice by psychologists and other practitioners. The Ohio Scales could be used to monitor change with clients with the confidence that this is currently the best routine outcome measure for young people available. Until further work is undertaken to collect New Zealand data (as described above), clinicians will need to use the measure in its current form and with its USA norms. Although not ideal, this is currently the case for many psychometric measures used by psychologists in New Zealand. Eventually it would be desirable for a wide range of measures to be validated here and the results of this research suggest towards which measures these efforts should be directed.

While the studies that make up this research contributed to the achievement of three specific aims, further implications for services wishing to introduce routine outcome measurement arose. These came from the responses to the survey and focus groups, which showed practitioners here had similar thoughts on the topic to those in other

countries. Practitioners' concerns about the use of outcome information should be heeded by services intending to implement routine outcome measurement in order to secure their willingness to use measures in their practice. This research indicated there is strong opposition to outcome data being used to allocate resources or funding, or to override the individualisation of intervention plans. This underscores the importance of communicating the purposes of outcome measurement to practitioners clearly. This should include ensuring they do not feel their professional autonomy is being threatened by the process and that measurement is seen as a valuable activity to improve effectiveness with individual clients and collectively as a service. This research showed that, like others overseas, New Zealand practitioners weight their professional opinion heavily in their evaluation of therapeutic progress, which is consistent with literature that practitioners value clinical judgment over psychometric evidence (Tansella, 2002). Psychometric information is a more reliable source of data than clinical judgment and so can be expected to increase the effectiveness of interventions by facilitating the timely identification of poor progress (Lambert, 2010). It would be important to emphasise that there are benefits of outcome measurement over and above accountability, which is why it is recommended as part of good clinical practice (Johnston & Gowers, 2005; Lambert & Hawkins, 2004).

Another strong theme from practitioners taking part in this research was for the information from measures to be integrated with other clinical information and this is also consistent with good practice. Multiple methods and sources should be used to gather information about client progress during interventions, and outcome measures are one method that enhances clinical decision-making (Hatfield & Ogles, 2006).

This research also revealed that limited knowledge about the measures available or having unsuitable measures to choose from were barriers to practitioners measuring outcome. There was a general lack of understanding of the notion of outcome measurement (and which measures were suited to this), and few participants reported having experience using them. This should include education about what measures can and cannot do, as many responses indicated absent or erroneous knowledge of psychometric assessment and interpretation. A recent article describing the experience of implementing routine outcome measurement in a psychology service in the UK, outlined how management input and training for clinicians helped lift self-report questionnaire completion rates at the initial assessment from 61% to 91% (Rao, et al., 2010). In addition, an Australian study found that practitioners attitudes to using feedback from outcome measures in their practice improved following a training

programme (Willis, Deane, & Coombs, 2009). Clearly, further training about outcome measures would be beneficial for HCN practitioners, and perhaps in other services wishing to implement a measurement protocol.

For services that already use routine outcome measurement, this research provides up-to-date information about the strengths of the outcome measures available for young people. The results could be used by these services to reconsider whether the measures they are currently using are the most suitable for their needs. For those who work from a scientist-practitioner perspective, such as New Zealand clinical psychologists (Evans & Fitzgerald, 2007), it is imperative that applied practice is continually adapted to reflect the latest research developments. Therefore, as the Ohio Scales has now been identified as the best routine outcome measure in two comprehensive reviews undertaken almost a decade apart, services using other measures should seriously consider it an alternative.

This research began with an investigation of current opinion and practices in outcome measurement for a specific multidisciplinary setting. However, this led to a wide-ranging search and review of potential outcome measures, and investigation of the Ohio Scales in particular, that provided results of interest to a broad range of services. Individual clinicians, private and public multidisciplinary services, and researchers can make use of the results of this research in a practical manner. As others have highlighted previously, there is a gap between measures used in treatment-outcome research and applied practice (Froyd, et al., 1996), and a degree of chaos has developed in the field (Hatfield & Ogles, 2004). This research bridges this gap by providing guidance to those selecting an outcome measure for routine use in multidisciplinary interventions. Lambert's (2010) call to resist "the temptation to go 'part-way'" (p. 253) in implementing routine measure use should be heeded by those wishing to replicate the benefits of measurement reported in the literature in real world services. Above all, the results support the continued validation and use of the Ohio Scales and this measure ought to be considered as a strong candidate for routine outcome measurement across New Zealand services for young people.

Knowing is not enough, we must apply. Willing is not enough, we must do.

– Goethe

REFERENCES

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington DC, USA: American Psychiatric Association.
- Amin, S., Singh, S. P., Croudace, T., Jones, P., Medley, I., & Harrison, G. (1999). Evaluating the Health of the Nation Outcome Scales: Reliability and validity in a three-year follow up of first-onset psychosis. *British Journal of Psychiatry*, *174*, 399-483.
- Andrews, G. (1999). Efficacy, effectiveness and efficiency in mental health service delivery. *Australian and New Zealand Journal of Psychiatry*, *33*, 316-322.
- Andrews, G., & Page, A. C. (2005). Outcome measurement, outcome management, and monitoring. *Australian and New Zealand Journal of Psychiatry*, *39*, 649-651.
- Andrews, G., Peters, L., & Teesson, M. (1994). *The measurement of consumer outcomes in mental health*. Canberra: Australian Government Publishing.
- Aylward, G. (1994). *Practitioner's guide to developmental and psychological testing*. USA: Springer.
- Barry, A., Baken, D., Fitzgerald, J., Evans, I., & Harvey, S. (2007). *Preparing a formal outcome monitoring protocol for use with High and Complex Needs cases*. Unpublished report for the High and Complex Needs Intersectoral Unit. Massey University/The Psychology Centre. Palmerston North/Hamilton, NZ.
- Beavers, R., & Hampson, R. B. (2000). The Beavers Systems Model of Family Functioning. *Journal of Family Therapy*, *22*, 128-143.

- Beavers, W. R., Hampson, R. B., & Hulgus, Y. F. (1985). Commentary: The Beavers systems approach to family assessment. *Family Process, 24*, 398-405.
- Bebbington, P., Brugha, T., Hill, T., Marsden, L., & Window, S. (1999). Validation of the Health of the Nation Outcome Scales. *British Journal of Psychiatry, 174*, 389-394.
- Bickman, L. (2008). Why don't we have effective mental health services? *Administration and Policy in Mental Health, 35*, 437-439.
- Bickman, L., Nurcombe, B., Townsend, C., Belle, M., Schut, J., & Karver, M. (1998). *Consumer measurement systems and child and adolescent mental health*. Canberra, Australia: Department of Health and Family Services.
- Bickman, L., Rosof-Williams, J., Salzer, M. S., Summerfelt, W. T., Noser, K., Wilson, S. J., et al. (2000). What information do clinicians value for monitoring adolescent client progress and outcomes? *Professional Psychology: Research and Practice, 31*(1), 70-74.
- Birleson, P. (1998). Learning organisations: A suitable model for improving mental health services. *Australian and New Zealand Journal of Psychiatry, 32*, 214-222.
- Black, S., Pulford, J., Christie, G., & Wheeler, A. (2010). Differences in New Zealand secondary school students' reported strengths and difficulties. *New Zealand Journal of Psychology, 39*(1), 19-23.
- Blisker, D., & Goldner, E. M. (2002). Routine outcome measurement by mental health-care providers: Is it worth doing? *The Lancet, 360*, 1689-1690.
- Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Thousand Oaks, USA: Sage.
- Boyd, C. P., Gullone, E., Needleman, G. L., & Burt, T. (1997). The Family Environment Scale: Reliability and normative data for an adolescent sample. *Family Process, 36*, 369-373.
- Brann, P., & Coleman, G. (2010). On the meaning of change in a clinician's routine measure of outcome: HoNOSCA. *Australian and New Zealand Journal of Psychiatry, 44*, 1097-1104.
- Brann, P., Coleman, G., & Luk, E. (2001). Routine outcome measurement in a child and adolescent mental health service: An evaluation of HoNOSCA. *Australian and New Zealand Journal of Psychiatry, 35*, 370-376.
- Burlingame, G. M., Mosier, J. I., Wells, M. G., Atkin, Q. G., Lambert, M., Whoolery, M. L., et al. (2001). Tracking the influence of mental health treatment: The

- development of the Youth Outcome Questionnaire. *Clinical Psychology and Psychotherapy*, 8, 361-379.
- Burlingame, G. M., Wells, M. G., & Lambert, M. J. (1996). *Youth Outcome Questionnaire*. Stevenson, USA: American Professional Credentialing Services.
- Bylund, C. (2008). Structured patient-clinician communication using DIALOG improves patient quality of life. *Evidence Based Mental Health*, 11(3), 89.
- Callaly, T., Coombs, T., & Berk, M. (2003). Routine outcome measurement by health-care providers. [Letter to the editor]. *The Lancet*, 361, 1137-1138.
- Calvert, S., & Lightfoot, S. (2002). Working with children with complex presentations: A New Zealand approach. *New Zealand Journal of Psychology*, 31, 65-72.
- Cameron, F. J., Smidst, D., Hesketh, K., Wake, M., & Northam, E. A. (2003). Early detection of the emotional and behavioural problems in children with diabetes: The validity of the Child Health Questionnaire as a screening instrument. *Diabetic Medicine*, 20, 646-650.
- Carr, A. (2006). *The handbook of child and adolescent clinical psychology: A contextual approach* (2nd ed.). London: Routledge.
- Church, J. (2003). *The definition, diagnosis and treatment of children and youth with severe behaviour difficulties: A review of the research*. Wellington, NZ: Ministry of Education.
- Ciarlo, J. A. (1982). Accountability revisited: The arrival of client outcome evaluation. *Evaluation and Program Planning*, 5, 31-36.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284-290.
- Code of Ethics Review Group. (2002). *Code of ethics for psychologists working in Aotearoa/New Zealand*. Wellington: New Zealand Psychologists Board.
- Conners, C. K. (1997). *Conners' Rating Scales - Revised: User's manual*. New York: Multi-Health Systems.
- Coolican, H. (1999). *Research methods and statistics in psychology* (3rd ed.). London, UK: Hodder & Stoughton.
- Coombs, T., & Meehan, T. (2003). Mental health outcomes in Australia: Issues for mental health services (Editorial). *International Journal of Mental Health Nursing*, 12, 163-164.
- Coombs, T., & Meehan, T. (2005). Standardized routine outcome measurement: Response to Lakeman. *International Journal of Mental Health Nursing*, 14, 215-217.

- Corcoran, K., & Vandiver, V. (1996). *Maneuvering the maze of managed care: Skills for mental health practitioners*. New York, USA: The Free Press.
- Cottrell, D., & Kraam, A. (2005). Growing up? A history of CAMHS (1987-2005). *Child and Adolescent Mental Health, 10*(3), 111-117.
- Cross, T. P., & Saxe, L. (1997). Many hands make mental health systems of care a reality: Lessons from the mental health services programs for youth. In C. T. Nixon & D. A. Northrup (Eds.), *Evaluating mental health services: How do programs for children "work" in the real world?* (pp. 45-72). Thousand Oaks, USA: Sage.
- de Klerk, G. (2008). Cross-cultural testing. In M. Born, C. D. Foxcroft & R. Butter (Eds.), *Online Readings in Testing and Assessment*: International Test Commission. Retrieved from <http://www.intestcom.org/Publications/ORTA.php>.
- De Los Reyes, A., & Kazdin, A. E. (2006). Informant discrepancies in assessing child dysfunction relate to dysfunction within mother-child interactions. *Journal of Child and Family Studies, 15*(5), 645-663.
- Denzin, N. K., & Lincoln, Y. S. (Eds.). (2005). *The Sage handbook of qualitative research* (3rd ed.). Thousand Oaks, USA: Sage.
- Dowell, K. A., & Ogles, B. M. (2008). The Ohio Scales Youth form: Expansion and validation of a self-report outcome measure for children. *Journal of Child and Family Studies, 17*, 291-305.
- Drotar, D., Schwartz, L., Palermo, T. M., & Burant, M. (2006). Factor structure of the Child Health Questionnaire - Parent form in pediatric populations. *Journal of Pediatric Psychology, 31*, 127-138.
- Dunn, T. W., Burlingame, G. M., Walbridge, M., Smith, J., & Crum, M. J. (2005). Outcome assessment for children and adolescents: Psychometric validation of the Youth Outcome Questionnaire 30.1 (Y-OQ-30.1). *Clinical Psychology and Psychotherapy, 12*, 388-401.
- Durie, M. (2001). *Mauri ora: The dynamics of Māori mental health*. Melbourne: Oxford University Press.
- Eatwell, J., & Wilson, I. (2007). The effective use of psychometric assessments in decision making. In I. Evans, J. Rucklidge & M. O'Driscoll (Eds.), *Professional practice of psychology in Aotearoa New Zealand* (pp. 317-333). Wellington: New Zealand Psychological Society.
- Eckert, P. A. (1994). Cost control through quality improvement: The new challenge for psychology. *Professional Psychology: Research and Practice, 25*(1), 3-8.

- Eggelston, M. J. F., & Watkins, W. G. A. (2008). Mental health services for children and adolescents in New Zealand, outcomes, and the Health of the Nation Outcome Scale for Children and Adolescents (HoNOSCA). *New Zealand Medical Journal*, 121(1271), 83-91.
- Eisen, S. V., Leff, H. S., & Schaefer, E. (1999). Implementing outcome systems: Lessons from a test of the BASIS-32 and the SF-36. *Journal of Behavioural Health Services and Research*, 26(1), 18-27.
- Epstein, N. B., Baldwin, L., & Bishop, D. S. (1983). The McMaster Family Assessment Device. *Journal of Marital and Family Therapy*, 9, 171-180.
- Evans, I., & Fitzgerald, J. (2007). Integrating research and practice in professional psychology: Models and paradigms. In I. Evans, J. Rucklidge & M. O'Driscoll (Eds.), *Professional practice of psychology in Aotearoa New Zealand* (pp. 283-300). Wellington: New Zealand Psychological Society.
- Eysenck, H. J. (1952). The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology*, 16, 319-324.
- Fitzgerald, J. (2007). Assessment of outcomes in mental health settings. In I. Evans, J. Rucklidge & M. O'Driscoll (Eds.), *Professional practice of psychology in Aotearoa New Zealand* (pp. 335-350). Wellington: New Zealand Psychological Society.
- Fitzgerald, J., Hodgetts, A., Ryan, J., Brassington, J., Collier, J., & Augustine, T. (2006). *A review of progress and outcome measures: Use with Sensitive Claims clients in Aotearoa/New Zealand*. Unpublished technical report for ACC. The Psychology Centre. Hamilton, NZ.
- Friedman, R. M. (1997). Services and service delivery systems for children with serious emotional disorders: Issues in assessing effectiveness. In C. T. Nixon & D. A. Northrup (Eds.), *Evaluating mental health services: How do programs for children "work" in the real world?* (pp. 16-44). Thousand Oaks, USA: Sage.
- Friesen, B. J., & Winters, N. C. (2003). The role of outcomes in systems of care: Quality improvement and program evaluation. In A. J. Pumariega & N. C. Winters (Eds.), *The handbook of child and adolescent systems of care: The new community psychiatry* (pp. 459-486). San Francisco, USA: Jossey-Bass.
- Froyd, J., Lambert, M. J., & Froyd, J. (1996). A review of practices of psychotherapy outcome measurement. *Journal of Mental Health*, 5(1), 11-16. Retrieved from doi:10.1080/09638239650037144

- Gaines, P., Bower, A., Buckingham, B., Eagar, K., Burgess, P., & Green, J. (2003). *New Zealand mental health classification and outcomes study: Final report*. Auckland, NZ: Health Research Council.
- Garland, A. F., Kruse, M., & Aarons, G. A. (2003). Clinicians and outcome measurement: What's the use? *Journal of Behavioural Health Services and Research, 30*(4), 393-405.
- Garralda, E., & Yates, P. (2000). HoNOSCA: Uses and limitations. *Child Psychology and Psychiatry Review, 5*(3), 131-132.
- Garralda, M. E., Yates, P., & Higgins, I. (2000). Child and adolescent mental health service use: HoNOS-CA as an outcome measure. *British Journal of Psychiatry, 177*, 52-58.
- Geisinger, K. (1994). Cross-cultural normative assessment: Translation and adaption issues influencing the normative interpretation of assessment instruments. *Psychological Assessment, 6*, 304-312.
- Gilbody, S., House, A., & Sheldon, T. (2002). Psychiatrists in the UK do not use outcome measures: National Survey. *British Journal of Psychiatry, 180*, 101-103.
- Gilgun, J. F. (1999). CASPARS: New tools for assessing client risks and strengths. *Families in Society, 80*, 450-459.
- Gilgun, J. F. (2004). Qualitative methods and the development of clinical assessment tools. *Qualitative Health Research, 14*, 1008-1014.
- Gilgun, J. F., Keskinen, S., Marti, D. J., & Rice, K. (1999). Clinical applications of the CASPARS instruments: Boys who act out sexually. *Families in Society, 80*, 629-641.
- Godfrey, H. P. D., Frost, S., Snelling, E., Knight, R. G., Shelton, E. J., & Longmore, B. E. (1986). AAMD Adaptive Behaviour Scale: Normative, reliability and validity data. *New Zealand Journal of Psychology, 15*, 62-67.
- Goldney, R. D., Fisher, L. J., & Walmsley, S. H. (1998). The Health of the Nation Outcome Scales in psychiatric hospitalisation: A multicentre study examining outcome and prediction of length of stay. *Australian and New Zealand Journal of Psychiatry, 32*, 199-205.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38*, 581-586.
- Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire (SDQ). *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 1337-1345.

- Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for psychiatric disorders in a community sample. *British Journal of Psychiatry, 177*, 534-539.
- Goodman, R., Meltzer, H., & Bailey, V. (2003). The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *International Review of Psychiatry, 15*, 173-177.
- Gordon, S., Ellis, P., Haggerty, C., Pere, L., Platz, G., & McLaren, K. (2004). *Preliminary work towards the development of a self-assessed measure of consumer outcome*. Auckland: Health Research Council of New Zealand.
- Gowers, S. G., Harrington, R. C., Whitton, A., Beevor, A., Lelliott, P., Jezzard, R., et al. (1999). Health of the Nation Outcome Scales for Children and Adolescents (HoNOS-CA): Glossary for HoNOS-CA score sheet. *British Journal of Psychiatry, 174*, 428-431.
- Gowers, S. G., Harrington, R. C., Whitton, A., Lelliott, P., Beevor, A., Wing, J., et al. (1999). Brief scale for measuring the outcomes of emotional and behavioural disorders in children. *British Journal of Psychiatry, 174*, 413-416.
- Gowers, S. G., Levine, W., Bailey-Rogers, S., Shore, A., & Burhouse, E. (2002). Use of a routine, self-report outcome measure (HoNOSCA-SR) in two adolescent mental health services. *British Journal of Psychiatry, 180*, 266-269.
- Groth-Marnat, G. (2003). *Handbook of psychological assessment* (4th ed.). New Jersey, USA: Wiley.
- Guthrie, D., McIntosh, M., Callaly, T., Trauer, T., & Coombs, T. (2008). Consumer attitudes towards the use of routine outcome measures in a public mental health service: A consumer-driven study. *International Journal of Mental Health Nursing, 17*, 92-97.
- Hambleton, R. (2001). The next generation of the ITC Test Translation and Adaptation Guidelines. *European Journal of Psychological Assessment, 17*, 164-172.
- Hampson, R. B., Hulgus, Y. F., & Beavers, W. R. (1991). Comparisons of self-report measures of the Beavers' Systems Model and Olson's Circumplex Model. *Journal of Family Psychology, 4*, 326-340.
- Hannan, C., Lambert, M. J., Harmon, C., Nielson, S., Smart, D. W., Shimokawa, K., et al. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology, 61*(2), 155-163.
- Hanssen-Bauer, K., Aalen, O. O., Ruud, T., & Heyerdahl, S. (2007). Inter-rater reliability of clinician-rated outcome measures in child and adolescent mental

- health services. *Administration and Policy in Mental Health*, 34, 504-512. doi: 10.1007/s10488-007-0134-y
- Hanssen-Bauer, K., Langsrud, O., Kvernmo, S., & Heyerdahl, S. (2010). Clinician-rated mental health in outpatient child and adolescent mental health services: Associations with parent, teacher and adolescent ratings. *Child and Adolescent Mental Health*, 4(29). Retrieved from <http://www.capmh.com/content/4/1/29> doi:10.1186/1753-2000-4-29
- Happell, B. (2008). Meaningful information or a bureaucratic exercise? Exploring the value of routine outcome measurement in mental health. *Issues in Mental Health Nursing*, 29, 1098-1114.
- Harmon, C., Hawkins, E. J., Lambert, M. J., Slade, K., & Whipple, J. L. (2005). Improving outcomes for poorly responding clients: The use of clinical support tools and feedback to clients. *Journal of Clinical Psychology/In session*, 61(2), 175-185.
- Harnett, P. H., Loxton, N. J., Sadler, T., Hides, L., & Baldwin, L. (2005). The Health of the Nation Outcome Scales for Children and Adolescents in an adolescent inpatient sample. *Australian and New Zealand Journal of Psychiatry*, 39, 129-135.
- Harvey, S., Barry, A., Fitzgerald, J., Evans, I., & Bennett, S. (2007). *Best practice outcome assessment for children and youth with high and complex needs*. Unpublished report for the High and Complex Needs Intersectoral Unit. Massey Psychology Clinic/The Psychology Centre. Palmerston North/Hamilton, NZ.
- Hatfield, D., McCullough, L., Frantz, S., & Krieger, K. (2009). Do we know when our clients get worse? An investigation of therapists' ability to detect negative client change. *Clinical Psychology and Psychotherapy*. Retrieved from www.interscience.wiley.com doi:10.1002/ccp
- Hatfield, D., & Ogles, B. M. (2004). The use of outcome measures by psychologists in clinical practice. *Professional Psychology: Research and Practice*, 35(5), 485-491.
- Hatfield, D., & Ogles, B. M. (2006). The influence of outcome measures in assessing client change and treatment decisions. *Journal of Clinical Psychology*, 62(3), 325-337.
- Hatfield, D., & Ogles, B. M. (2007). Why some clinicians use outcome measures and others do not. *Administration and Policy in Mental Health and Mental Health Services Research*, 34, 283-291.

- Hawkins, R. P., Almeida, M. C., Fabry, B., & Reitz, A. L. (1992). A scale to measure restrictiveness of living environments for troubled children and youth. *Hospital and Community Psychiatry, 43*(1), 54-58.
- Hazelton, M., & Farrell, G. A. (1998). *Evaluating the outcomes of mental health care: An introduction*. Greenacres, Australia: Australian and New Zealand College of Mental Health Nurses.
- Healy, J. (2007). *The Ohio Consumer Outcomes Initiative*. Paper presented at the He Kakano: Australasian Mental Health Outcomes Conference, Wellington.
- Healy, J., & Tam, K. K. (2009). *Ohio Mental Health Consumer Outcomes System Report 20: Early warning system for identifying youth at risk of treatment failure*. Retrieved from <http://www.mh.state.oh.us/oper/outcomes>
- High and Complex Needs Intersectoral Unit. (2005). *Request for tender for research project: Report on possible outcome measures for individualised plans for children and young people with high and complex needs*. Wellington, NZ: High and Complex Needs Intersectoral Unit.
- High and Complex Needs Intersectoral Unit. (n.d.). *Intersectoral strategy for children and young people with high and complex needs*. Retrieved from <http://www.hcn.govt.nz>
- Hill, C. E., & Lambert, M. J. (2004). Methodological issues in studying psychotherapy processes and outcome. In M. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behaviour change* (5th ed., pp. 84-135). New York, USA: John Wiley & Sons.
- Hodges, K. (2004). Using assessment in everyday practice for the benefit of families and practitioners. *Professional Psychology: Research and Practice, 35*(5), 449-456.
- Hodgetts, A., Fitzgerald, J., Brassington, J., Ryan, J., Collier, J., & Augustine, T. (2005). *Measuring outcome: A survey of ACC Sensitive Claim practitioners*. Unpublished technical report for ACC. Massey University. Palmerston North, NZ.
- Hudson, W. W. (1982). *The clinical measurement package: A field manual*. Tallahassee, USA: WALMYR Publishing.
- Hudson, W. W. (1990a). *Index of Family Relations*. Tempe, USA: WALMYR Publishing.
- Hudson, W. W. (1990b). *The WALMYR Assessment Scales scoring manual*. Tempe, USA: WALMYR Publishing.

- Hunter, J., Higginson, I., & Garralda, E. (1996). Systematic literature review: Outcome measures for child and adolescent mental health services. *Journal of Public Health Medicine, 18*(2), 197-206.
- Huxley, P. (1994). Outcome measurement in work with children: Comparing plans in the UK with experience in the US. *Child Abuse Review, 3*, 120-133.
- Jaffa, T. (2000). HoNOSCA: Is the enthusiasm justified? *Child Psychology and Psychiatry Review, 5*(3), 130.
- Jellinek, M. S., Murphy, J. M., Little, M., Pagano, M. E., Comer, D., & Kelleher, K. (1999). The use of the Pediatric Symptom Checklist to screen for psychosocial problems in pediatric primary care: A national feasibility study. *Archives of Pediatric and Adolescent Medicine, 153*, 254-260.
- Jensen-Doss, A., & Hawley, K. M. (2010). Understanding barriers to evidence-based assessment: Clinician attitudes toward standardized assessment tools. *Journal of Clinical Child and Adolescent Psychology, 39*(6), 885-896.
- Johnson, M. A., Stone, S., Lou, C., Vu, C., Ling, J., Mizrahi, P., et al. (2006). *Family assessment in child welfare services: Instrument comparisons*. USA: Centre for Social Services Research, University of California at Berkley.
- Johnston, C., & Gowers, S. G. (2005). Routine outcome measurement: A survey of UK child and adolescent mental health services. *Child and Adolescent Mental Health, 10*(3), 133-139.
- Kabakoff, R. I., Miller, I. W., Bishop, D. S., Epstein, N. B., & Keitner, G. I. (1990). A psychometric study of the McMaster Family Assessment Device in psychiatric, medical and non-clinical samples. *Journal of Family Psychology, 3*, 431-439.
- Kamberelis, G., & Dimitriadis, G. (2005). Strategic articulations of pedagogy, politics, and inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 887-907). Thousand Oaks, USA: Sage.
- Kazdin, A. E. (2004). Psychotherapy for children and adolescents. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behaviour change* (5th ed., pp. 543-589). New York, USA: John Wiley & Sons.
- Kendall, P. C., & Sheldrick, R. C. (2000). Normative data for normative comparisons. *Journal of Consulting and Clinical Psychology, 68*(5), 767-773.
- Kirk, R., & Reed-Ashcraft, K. (1998a). *Definitions for the NCFAS: North Carolina Family Assessment Scale for Intensive Family Preservation Services programs, Version 2.0* Retrieved from <http://www.nfnpn.org/preservation>
- Kirk, R., & Reed-Ashcraft, K. (1998b). *NCFAS: North Carolina Family Assessment Scale, Version 2.0* Retrieved from <http://www.nfnpn.org/preservation>

- Kirk, R., & Reed-Ashcraft, K. (n.d.-a). *NCFAS, North Carolina Family Assessment Scale: Research report* Retrieved from <http://www.nfnp.org/preservation>
- Kirk, R., & Reed-Ashcraft, K. (n.d.-b). *User's guide for the North Carolina Family Assessment Scale (NCFAS), Version 2.0* Retrieved from <http://www.nfnp.org/preservation>
- Koch, J. R., Lewis, A., & McCall, D. (1998). A multistakeholder-driven model for developing an outcome management system. *Journal of Behavioural Health Services and Research, 25*(2), 151-162.
- Kriebler, T. (2003). Towards an outcome-based mental health policy for New Zealand. *Australasian Psychiatry, 11*(Supplement), 78-82.
- Krueger, R. A., & Casey, M. A. (2000). *Focus groups: A practical guide for applied research* (3rd ed.). Thousand Oaks, USA: Sage.
- Lakeman, R. (2004). Standardised routine outcome measurement: Pot holes in the road to recovery. *International Journal of Mental Health Nursing, 13*, 210-215.
- Lambert, M. J. (2010). *Prevention of treatment failure: The use of measuring, monitoring, and feedback in clinical practice*. Washington DC: American Psychological Association.
- Lambert, M. J., Bergin, A. E., & Garfield, S. L. (2004). Introduction and historical overview. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behaviour change* (5th ed., pp. 3-15). New York, USA: John Wiley & Sons.
- Lambert, M. J., Harmon, C., Slade, K., Whipple, J. L., & Hawkins, E. J. (2005). Providing feedback to psychotherapists on their patients' progress: Clinical results and practice suggestions. *Journal of Clinical Psychology, 61*(2), 165-174.
- Lambert, M. J., & Hawkins, E. J. (2004). Measuring outcome in professional practice: Considerations in selecting and using brief outcome instruments. *Professional Psychology: Research and Practice, 35*(5), 492-499.
- Lambert, M. J., & Ogles, B. M. (2004). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behaviour change* (5th ed., pp. 139-193). New York, USA: John Wiley & Sons.
- Lambert, M. J., Okiijsji, J., Finch, A., & Johnson, L. (1998). Outcome assessment: From conceptualization to implementation. *Professional Psychology: Research and Practice, 29*(1), 63-70.

- Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy, 48*(1), 72-79.
- Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003). Is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clinical Psychology: Science and Practice, 10*(3), 288-301.
- Lambert, M. J., Whipple, J. L., Smart, D. W., Vermeersch, D. A., Nielsen, S. L., & Hawkins, E. J. (2001). The effects of providing therapists with feedback on patient progress during psychotherapy: Are outcomes enhanced? *Psychotherapy Research, 11*, 49-68.
- Lambert, M. J., Whipple, J. L., Vermeersch, D. A., Smart, D. W., Hawkins, E. J., Nielsen, S. L., et al. (2002). Enhancing psychotherapy outcomes via providing feedback on client progress: A replication. *Clinical Psychology and Psychotherapy, 9*, 91-103.
- Lambert, W., Salzer, M. S., & Bickman, L. (1998). Clinical outcome, consumer satisfaction, and ad hoc ratings of improvement in children's mental health. *Journal of Consulting and Clinical Psychology, 66*(2), 270-279.
- Lourie, I. S. (2003). A history of community child mental health. In A. J. Pumariega & N. C. Winters (Eds.), *The handbook of child and adolescent systems of care: The new community psychiatry* (pp. 1-16). San Francisco, USA: Jossey-Bass.
- Lyons, J. S., Howard, K. I., O'Mahoney, M. T., & Lish, J. D. (1997). *The measurement and management of clinical outcomes in mental health*. New York, USA: John Wiley & Sons.
- Margison, F. R., Barkham, M., Evans, C., McGrath, G., Clark, J. M., Audin, K., et al. (2000). Measurement and psychotherapy: Evidence-based practice and practice-based evidence. *British Journal of Psychiatry, 177*, 123-130.
- McDaniel, N. (1996). Key stages and common themes in outcome measurement. *Evaluation Exchange, 2*(3). Retrieved from Harvard Family Research Project website: <http://www.hfrp.org/evaluation/the-evaluation-exchange/issue-archive>
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.
- Melendez, G. (2003). Effects of feedback to the therapist on child clinical outcomes: A patient-oriented approach to demonstrating effective practice. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 63*(8-B), 3930.
- Merry, S., Stasiak, K., Parkin, A., Seymour, F., Lambie, I., Crengle, S., et al. (2004). *Child and youth outcome measures: Examining current use and acceptability of*

- measures in mental health services and recommending future directions.* Auckland, NZ: Health Research Council.
- Miller, I. W., Epstein, N. B., Bishop, D. S., & Keitner, G. I. (1985). The McMaster Family Assessment Device: Reliability and validity. *Journal of Marital and Family Therapy, 11*, 345-356.
- Miller, S., Duncan, B., Brown, J., Sorrell, R., & Chalk, M. (2006). Using formal client feedback to improve retention and outcome: Making ongoing, real-time assessment feasible. *Journal of Brief Therapy, 5*(1), 5-22.
- Miller, S., Duncan, B., Sorrell, R., & Brown, G. (2005). The Partners for Change Outcome Management System. *Journal of Clinical Psychology/In session, 61*(2), 199-208.
- Ministry of Health. (2005). *National mental health information strategy 2005-2010.* Wellington, NZ: Ministry of Health.
- Moos, R., & Moos, B. (1994). *Family Environment Scale manual: Development, applications, research* (3rd ed.). Palo Alto, USA: Consulting Psychologist Press.
- Moos, R. H. (1990). Conceptual and empirical approaches to developing family-based assessment procedures: Resolving the case of the Family Environment Scale. *Family Process, 29*, 199-208.
- Myers, D. G. (2009). *Exploring social psychology* (5th ed.). New York: McGraw Hill.
- Newnham, E., & Page, A. C. (2007). Client-focused research: New directions in outcome assessment. *Behaviour Change, 24*(1), 1-6.
- Niumata-Faleafa, M., & Lui, D. (2005). *A preliminary report on outcome measures for Pacific Island peoples.* Auckland: Te Pou.
- Nixon, C. T. (1997). How does evaluation of mental health services for children work in the real world? In C. T. Nixon & D. A. Northrup (Eds.), *Evaluating mental health services: How do programs for children "work" in the real world?* (pp. 1-15). Thousand Oaks, USA: Sage.
- Noser, K., & Bickman, L. (2000). Quality indicators of children's mental health services: Do they predict improved client outcomes? *Journal of Emotional and Behavioral Disorders, 8*(1), 9-18, 26.
- Ogles, B. M. (2009). *Handouts to accompany Ohio Scales training video.* Retrieved from <http://sites.google.com/site/ohioscales/training-aids/helps>
- Ogles, B. M., Melendez, G., Davis, D. C., & Lunnen, K. M. (1999). *The Ohio Youth Problems, Functioning, and Satisfaction Scales (Short form): User's manual* Retrieved from <http://www.mh.state.oh.us/oper/outcomes>

- Ogles, B. M., Melendez, G., Davis, D. C., & Lunnen, K. M. (2000). *The Ohio Youth Problems, Functioning, and Satisfaction Scales: Technical manual*. Retrieved from <http://www.mh.state.oh.us/oper/outcomes>
- Ogles, B. M., Melendez, G., Davis, D. C., & Lunnen, K. M. (2001). The Ohio Scales: Practical outcome assessment. *Journal of Child and Family Studies, 10*, 199-212.
- Ogles, B. M., & Williams, C. (2009). *Using the Ohio Scales for family engagement, treatment planning, and outcomes tracking*. [Online video presentation and PowerPoint notes] Retrieved from <http://sites.google.com/site/ohioscales/training-aids/helps>
- Ollendick, T. H., Grills, A. E., & King, N. J. (2001). Applying developmental theory to the assessment and treatment of childhood disorders: Does it make a difference? *Clinical Psychology and Psychotherapy, 8*, 304-314.
- Orlinsky, D. E., Ronnestad, M. H., & Willutzki, U. (2004). Fifty years of psychotherapy process-outcome research: Continuity and change. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behaviour change* (5th ed., pp. 307-389). New York, USA: John Wiley & Sons.
- Orrell, M., Yard, P., Handysides, J., & Schapira, R. (1999). Validity and reliability of the Health of the Nation Outcome Scales in psychiatric patients in the community. *British Journal of Psychiatry, 174*, 409-412.
- Pagano, M. E., Cassidy, L. J., Little, M., Murphy, J. M., & Jellinek, M. S. (2000). Identifying psychosocial dysfunction in school-age children: The Pediatric Symptom Checklist as a self-report measure. *Psychology in the Schools, 37*, 91-106.
- Phelan, M., Wykes, T., & Goldman, H. (2001). Global functioning scales. In M. Tansella & G. Thornicroft (Eds.), *Mental health outcome measures* (2nd ed., pp. 19-33). London: Royal College of Psychiatrists.
- Phinney, J. (1992). The Multigroup Ethnic Identity Measure: A new scale for use with adolescents and young adults from diverse groups. *Journal of Adolescent Research, 7*, 156-176.
- Plante, T., Couchman, C., & Hoffman, C. (1998). Measuring treatment outcome and client satisfaction among children and families: A case report. *Professional Psychology: Research and Practice, 29*(1), 52-55.
- Priebe, S., McCabe, R., Bullenkamp, J., Hansson, L., Lauber, C., Martinez-Leal, R., et al. (2007). Structured patient-clinician communication and 1-year outcome in

- community mental healthcare: Cluster randomised controlled trial. *British Journal of Psychiatry*, 191, 420-426.
- Raat, H., Botterweck, A. M., Landgraf, J. M., Hoogeveen, W. C., & Essink-Bot, M. (2005). Reliability and validity of the short form of the Child Health Questionnaire for parents (CHQ-PF28) in large random school-based and general population samples. *Journal of Epidemiological Community Health*, 59, 75-82.
- Rao, A. S., Hendry, G., & Watson, R. (2010). The implementation of routine outcome measures in a Tier 3 Psychological Therapy Service: The process of enhancing data quality and reflections of implementation challenges. *Counselling and Psychotherapy Research*, 10(1), 32-38.
- Ridenour, T. Y. A., Daley, J., & Reich, W. (1999). Factor analyses of the Family Assessment Device. *Family Process*, 38, 497-510.
- Ridge, N. W., Warren, J. S., Burlingame, G. M., Wells, M. G., & Tumblin, K. M. (2009). Reliability and validity of the Youth Outcome Questionnaire self-report. *Journal of Clinical Psychology*, 65(10), 1115-1126.
- Robert, C., Lee, W., & Chan, K. (2006). An empirical analysis of measurement equivalence with the INDCOL measure of individualism and collectivism: Implications for valid cross-cultural inference. *Personnel Psychology*, 59, 65-99.
- Roberts, R., Phinney, J., Masse, L., Chen, Y., Roberts, C., & Romero, A. (1999). The structure of ethnic identity in young adolescents from diverse ethnocultural groups. *Journal of Early Adolescence*, 19, 301-322.
- Rock, D., Combrinck, J., & Groves, A. (2001). Issues associated with the implementation of routine outcome measures in public mental health services. *Australasian Psychiatry*, 9(1), 43-46.
- Rog, D. J., & Bickman, L. (1984). The feedback research approach to evaluation: A method to increase evaluation utility. *Evaluation and Program Planning*, 7, 169-175.
- Ruggeri, M. (2001). Measuring satisfaction with psychiatric services: Towards a multi-dimensional, multi-axial assessment of outcome. In M. Tansella & G. Thornicroft (Eds.), *Mental health outcome measures* (2nd ed., pp. 34-47). London: Royal College of Psychiatrists.
- Sapyta, J., Riemer, M., & Bickman, L. (2005). Feedback to clinicians: Theory, research, and practice. *Journal of Clinical Psychology/In session*, 61(2), 145-153.
- Sharma, V. K., Wilkinson, G., & Fear, S. (1999). Health of the Nation Outcome Scales: A case study in general psychiatry. *British Journal of Psychiatry*, 174, 395-398.

- Sines, J. O. (1970). Actuarial versus clinical prediction in psychopathology. *British Journal of Psychiatry*, 116, 129-144.
- Slade, M. (2002). Routine outcome assessment in mental health services. [Editorial]. *Psychological Medicine*, 32, 1339-1343.
- Slade, M., Beck, A., Bindman, J., Thornicroft, G., & Wright, S. (1999). Routine clinical outcome measures for patients with severe mental illness: CANAS and HoNOS. *British Journal of Psychiatry*, 174, 484-488.
- Slonim-Nevo, V., & Anson, Y. (1998). Evaluating practice: Does it improve treatment outcome? *Social Work Research*, 22(2), 66-74. Retrieved from Health Business Fulltext Elite database
- Sonnangburg, K. (1996). Meaningful measurement in psychotherapy. *Psychotherapy*, 33(2), 160-170.
- Speer, D. C. (1998). *Mental health outcome evaluation*. San Diego, USA: Academic Press.
- Spence, S., Donald, M., Dower, J., Woodward, R., & Lacherez, P. (2002). *Outcomes and indicators, measurement tools and databases for the National Action Plan for Promotion and Early Intervention for Mental Health 2000*. Retrieved from <http://www.mentalhealth.gov.au>
- Starfield, B., Bergner, M., Ensminger, M., Riley, A., Ryan, S., Green, B., et al. (1993). Adolescent health status measurement: Development of the Child Health and Illness Profile. *Pediatrics*, 91, 430-435.
- Stein, G. S. (1999). Usefulness of the Health of the Nation Outcome Scales. [Editorial]. *British Journal of Psychiatry*, 174, 375-377.
- Stewart, M. (2007). *Making HoNOS-CA useful in the MDT setting*. Paper presented at the He Kakano: Australasian Mental Health Outcomes Conference, Wellington.
- Sung, L., Greenberg, M. L., Doyle, J. J., Young, N. L., Ingber, S., Rubenstein, J., et al. (2003). Construct validation of the Health Utilities Index and the Child Health Questionnaire in children undergoing cancer chemotherapy. *British Journal of Cancer*, 88, 1185-1190.
- Tam, K. K., & Healy, J. (2007). *Ohio Mental Health Consumer Outcomes System Report 16: Benchmarks for reliable change and clinical significance on the Ohio Scales for youth - Problem Severity and Functioning scales*. Retrieved from <http://www.mh.state.oh.us/oper/outcomes>
- Tansella, M. (2002). The scientific evaluation of mental health treatments: An historical perspective. *Evidence Based Mental Health*, 5, 4-5.

- Te Pou. (2009). *Past, present and future: Vision paper*. Hamilton, NZ: Te Pou o Te Whakaaro Nui.
- Texas Department of Mental Health and Mental Retardation. (n.d.). *Validation and Norms for the Ohio Scales Among Children Served by the Texas Department of Mental Health and Mental Retardation*. Retrieved from www.dshs.state.tx.us/mhprograms
- The Werry Centre. (2010). *Evidence-based age-appropriate interventions: A guide for Child and Adolescent Mental Health Services (CAMHS)* (2nd ed.). Auckland: The Werry Centre for Child and Adolescent Mental Health Workforce Development.
- Trauer, T. (1998). Issues in the assessment of outcome in mental health. *Australian and New Zealand Journal of Psychiatry*, 32, 337-343.
- Trauer, T. (2003). Routine outcome measurement by health-care providers. [Letter to the editor]. *The Lancet*, 361, 1137.
- Trauer, T. (2010). Outcome measurement in chronic mental illness. *International Review of Psychiatry*, 22(2), 99-113.
- Trauer, T., Eagar, K., Gaines, P., & Bower, A. (2004). *New Zealand mental health consumers and their outcomes*. Auckland, NZ: Health Research Council.
- Tryon, G. S., & Winograd, G. (2011). Goal consensus and collaboration. *Psychotherapy*, 48(1), 50-57.
- Turchik, J. A., Karpenko, V., & Ogles, B. M. (2007). Further evidence of the utility and validity of a measure of outcome for children and adolescents. *Journal of Emotional and Behavioral Disorders*, 15(2), 119-128.
- Van de Vijver, F., & Hambleton, R. (1996). Translating tests: Some practical guidelines. *European Psychologist*, 1, 89-99.
- van Widenfelt, B., Treffers, P., de Beurs, E., Siebelink, B., & Koudijs, E. (2005). Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families. *Clinical Child and Family Psychology Review*, 8, 135-147.
- Wall, J. R., Busch, M., Koch, S. M., Alexander, G., Minnich, H., & Jackson-Walker, S. (2005). Accountability in child welfare services: Developing a statewide outcome evaluation program. *Psychological Services*, 2(1), 39-53.
- Waters, E. B., Salmon, L. A., & Wake, M. (2000). The parent form Child Health Questionnaire in Australia: Comparison of reliability, validity, structure and norms. *Journal of Pediatric Psychology*, 25, 381-391.

- Waters, E. B., Salmon, L. A., Wake, M., Wright, M., & Hesketh, K. (2001). The health and well-being of adolescents: A school-based population study of the self-report Child Health Questionnaire. *Journal of Adolescent Health, 29*, 140-149.
- Wells, M. G., Burlingame, G. M., Lambert, M. J., Hoag, M. J., & Hope, C. A. (1996). Conceptualization and measurement of patient change during psychotherapy: Development of the Outcome Questionnaire and Youth Outcome Questionnaire. *Psychotherapy: Theory, Research, Practice, Training, 33*, 275-283.
- Whipple, J. L., Lambert, M. J., Vermeersch, D. A., Smart, D. W., Nielsen, S. L., & Hawkins, E. J. (2003). Improving the effects of psychotherapy: The use of early identification of treatment failure and problem-solving strategies in routine practice. *Journal of Counseling Psychology, 50*(1), 59-68.
- Whoolery, M. L. (1997). *The test-retest reliability and internal consistency of the Youth Outcome Questionnaire*. Unpublished Masters thesis, Brigham Young University, Utah.
- Willis, A., Deane, F. P., & Coombs, T. (2009). Improving clinicians' attitudes towards providing feedback on routine outcome assessments. *International Journal of Mental Health Nursing, 18*, 211-215.
- Wing, J., Beevor, A., Curtis, R. H., Park, S. B. G., Hadden, S., & Burns, A. (1998). Health of the Nation Outcome Scales (HoNOS): Research and development. *British Journal of Psychiatry, 172*, 11-18.
- Yates, P., Garralda, M. E., & Higgins, I. (1999). Paddington Complexity Scale and Health of the Nation Outcome Scales for Children and Adolescents. *British Journal of Psychiatry, 174*, 417-423.
- Yates, P., Kramer, T., & Garralda, M. E. (2006). Use of a routine mental health measure in an adolescent secure unit. *British Journal of Psychiatry, 188*, 583-584.
- Young, W., Earle, J., & Dadebo, M. (2004). An intersectoral response to children with complex healthcare needs. *Canadian Journal of Educational Administration and Policy, 29*. Retrieved from <http://www.umanitoba.ca/publications/cjeap/articles/>
- Zimmerman, D. P. (1996). A comparison of commonly used treatment measures. In S. I. Pfeiffer (Ed.), *Outcome assessment in residential treatment* (pp. 49-69). New York: Haworth Press.

APPENDIX A – STUDY ONE: HCN PRACTITIONER SURVEY

- Information Sheet
- Survey



Survey of Outcome Measurement Practices used by Front-Line Staff with High and Complex Needs Children and Young People

INFORMATION SHEET

Introduction

We are conducting a survey as part of a broader research project entitled *Possible Outcome Measures for Individualised Plans for Children and Young People with High and Complex Needs*. Your experience and expertise is highly valued. We would very much appreciate hearing your views and invite you to take part in this project.

This particular phase of our project involves distributing a survey to front-line staff of High and Complex Needs Exceptions Funded clients (HCN Clients).

What is this project about?

With this survey, we hope to gather information about the outcome measurement practices that front-line staff use with HCN clients, in particular, the ways in which you are currently measuring change. We are an independent research team based at Massey University, Palmerston North, and the Psychology Centre in Hamilton. We are being funded by the High and Complex Needs Unit to conduct this research.

The purpose of the research is to identify outcome measurement tools and assessment practices that will aid front-line workers across agencies to track change for HCN clients they are working with.

An outcome measure is defined as a standardised format for collecting information about clients' progress and outcome. It may or may not be norm based. Examples of outcome measures include questionnaires, rating scales, checklists, and structured interviews or observations. Common measures amongst others include the Child Behaviour Checklist, HoNOSCA, the Kessler Screening Tool, CAGE, ASTTLE, PAT.

Am I eligible to take part?

You are eligible to take part in this study if you have previously been, or are currently part of an intersectoral team who is working with a High and Complex Needs Exceptions Funded client.

We hope to get as many people's views as possible, as this survey will help to guide our recommendations about outcome measurement approaches with HCN funded clients.

What am I being asked to do?

If you agree to take part, we'd like you to fill out a survey about your experience with HCN clients. The survey contains questions about your background and professional training, your experience with HCN clients, and your thoughts on outcome assessment and the monitoring of change over time. The survey is likely to take less than 45 minutes to complete.

- Continued, please turn over -

- In Confidence -

We don't anticipate that there would be any risk or discomfort to you as a result of participation.

What will happen to my information?

All returned surveys will be stored in a locked cabinet in the Psychology Clinic at Massey University. Be assured that no one will be able to identify you. Analysis of the survey data will be conducted by our research team in Palmerston North and Hamilton. At the end of the project, the surveys will be destroyed. We will send out a summary of our findings at the end of the study to all those who took part.

Information gathered from this survey will inform our recommendations to the HCN Intersectoral Unit on assessment approaches and measures for monitoring outcomes with High and Complex Needs Exceptions Funded clients.

What can I expect from the researchers?

If you decide to take part in this project, the researchers will respect your right to:

- ask any questions of the researchers about the study at any time during participation;
- decline to answer any particular question in the survey;
- withdraw from the survey;
- provide information on the understanding that it is completely confidential to the researchers;
- be given a summary of the findings.

Who can I speak with about taking part in this project?

If you have further questions or concerns, please do not hesitate to contact Shane Harvey or Amber Barry. Their contact details are below:

School of Psychology
Massey University
Private Bag 11-222
Palmerston North

Telephone: 06 356 9099 xtn 7171

Email: s.t.harvey@massey.ac.nz

We appreciate that you have a busy schedule and we are very grateful for your making time to consider completing this survey. We look forward to hearing your ideas.

The HCN Outcome Measurement Project Team

Shane Harvey	Ian Evans
Amber Barry	John Fitzgerald
Simon Bennett	Don Baken
Sarah Calvert	

This project has been reviewed and approved by the Massey University Human Ethics Committee, PN Application 06/40. If you have any concerns about the conduct of this research, please contact Professor Sylvia V Rumball, Chair, Massey University Campus Human Ethics Committee: Palmerston North, telephone 06 350 5249, email humanethicspn@massey.ac.nz.

- Continued, please turn over -

- In Confidence -



SURVEY OF OUTCOME MEASUREMENT USE

This survey is about the forms of outcome measurement you use and prefer with High and Complex Needs (HCN) funded clients

SECTION A: OUTCOME MEASUREMENT

1. Please complete the following:	
Discipline/Current Professional Title (e.g., Psychiatrist, Psychologist, Supervising Social Worker, Specialist Education Advisor etc.)	
Ethnicity	
Qualifications	
Professional association (e.g., ANZASW)	
2. How long have you been working in this professional capacity?	
3. How long have you been working with cases that require High and Complex Needs funding?	
4. How many HCN clients have you seen in total ?	
5. How long have you been working with HCN clients? (Months/Years)	
6. Please tell us your practice location(s) -	
Where is your current workplace? Please include region (e.g., Northland, Waikato) and nearest town or city (e.g., Masterton, Christchurch)	
Which organisation (e.g., Child Youth and Family Service, Ministry of Education, Special Education (GSE), District Health Board, Other Health agencies)?	

An outcome measure is defined as a *standardised format for collecting information about clients' progress and outcome*. It may or may not be norm based. Examples of outcome measures include questionnaires, rating scales, checklists, and structured interviews or observations. Common measures include the Child Behavior Checklist, HoNOSCA, Kessler Screening Tool, CAGE, ASTTLE, PAT.

7. Below are three groupings of HCN client characteristics. Please estimate the number for each grouping that applies to the HCN clients you have been involved with.

<i>Gender</i>	<input type="text"/>	Males
	<input type="text"/>	Females
<i>Age(s)</i>	<input type="text"/>	Children (aged between 0 and 10 years)
	<input type="text"/>	Adolescents (11 years to 15 years)
	<input type="text"/>	Adolescents (16 years to 21 years)
<i>Ethnicity</i>	<input type="text"/>	Pākehā
	<input type="text"/>	Māori
	<input type="text"/>	Pacific Island
	<input type="text"/>	Other (<i>Please state</i>) _____

8. Rate the statements below (using a 1—5 scale) according to how significant some of the following difficulties are for the HCN client(s) you are dealing with.
1 = not significant 2 3 = moderately significant 4 5 = most significant

1. Physical Disabilities & Health	<input type="text"/>	2. Mental Health	<input type="text"/>
3. Intellectual/ Developmental e.g., Autism	<input type="text"/>	4. Anti-social, Aggression	<input type="text"/>
5. Communication/ Speech	<input type="text"/>	6. Internalising e.g., withdrawn	<input type="text"/>
7. Learning, Achievement	<input type="text"/>	8. Lack of Cultural Identity	<input type="text"/>
9. Family/Whanau Dysfunction	<input type="text"/>	10. Lack of Attachment, Belonging	<input type="text"/>
11. Head injury/ Cognitive Difficulties	<input type="text"/>	12. Care and Protection	<input type="text"/>
13. Other (please state):	<input type="text"/>		<input type="text"/>

9. Describe your experience of using outcome measures with HCN clients

10. What groundwork needs to occur first with HCN clients before administering outcome measures?

11. Rate the statements below (using a 1—5 scale) according to how important each is for you in determining how your client is making progress.

1 = least significant 2 3 = moderately significant 4 5 = most significant

- _____ What the client tells me
- _____ What the client does (both during and between times you see them)
- _____ My observations (e.g., verbal and nonverbal cues/behaviours)
- _____ What others tell me (client's parent/caregiver, teacher, peers, siblings, etc.)
- _____ My own professional impression, based on training (e.g., training involves academic or professional courses, supervised practice)
- _____ My own professional impression, based on experience
- _____ Client responses to questionnaires/surveys (e.g., HoNOSCA, Child Behaviour Checklist)
- _____ Client responses to other monitoring measures (e.g., CAGE or Kessler from the Towards Well-Being System, Visual Analogue Scales, Subjective Units of Distress etc)
- _____ Client responses to satisfaction indicators (e.g., client feedback)
- _____ Client achieving set goals
- _____ File reports
- _____ Symptom (problem) reduction
- _____ Time spent in receipt of services (e.g., predetermined number of hours/sessions)
- _____ Other indicators

12. Rate the statements below (using a 1—5 scale) according to how important each is for guiding your recommendation that a High and Complex Needs client is ready to transition to mainstream services.

1 = least significant 2 3 = moderately significant 4 5 = most significant

- _____ What the client tells me
- _____ What the client does (both during and between times you see them)
- _____ My observations (e.g., verbal and nonverbal cues/behaviours)
- _____ What others tell me (client's parent/caregiver, teacher, peers, siblings, etc.)
- _____ My own professional impression, based on training (e.g., training involves academic or professional courses, supervised practice)
- _____ My own professional impression, based on experience
- _____ Client responses to questionnaires/surveys (e.g., HoNOSCA, Child Behaviour Checklist)
- _____ Client responses to other monitoring measures (e.g., CAGE or Kessler from the Towards Well-Being System, Visual Analogue Scales, Subjective Units of Distress etc)
- _____ Client responses to satisfaction indicators (e.g., client feedback)
- _____ Client achieving set goals
- _____ File reports
- _____ Symptom (problem) reduction
- _____ Time spent in receipt of services (e.g., predetermined number of hours/sessions)
- _____ Other indicators

13. Please list the outcome measures that you recommend could be used with HCN funded clients. Please answer the following questions for as many measures as you wish to comment on. If you wish to comment on more, make additional copies of this page.

Full name of the outcome measure/tool (e.g., The Kessler Screening Tool)

What is the extent of your experience with this measure (e.g., HCN clients, other clients, training)?

How long does it take to complete (minutes)?

How often should you use this outcome measure with HCN clients (please tick) and why?

Routinely	<input type="checkbox"/>	Why?
Occasionally	<input type="checkbox"/>	
Know about but don't use	<input type="checkbox"/>	

What are the advantages and strengths of this outcome measure/tool?

What are the limitations and weaknesses of this outcome measure/tool?

What domain(s) would you use this outcome measure/tool to assess (please tick)?

1. Physical Health	<input type="checkbox"/>	2. Mental Wellbeing	<input type="checkbox"/>
3. Development	<input type="checkbox"/>	4. Independence	<input type="checkbox"/>
5. Social Interaction	<input type="checkbox"/>	6. Communication/Speech	<input type="checkbox"/>
7. Educational (learning, achievement)	<input type="checkbox"/>	8. Cultural Identity	<input type="checkbox"/>
9. Family/Whanau Relationships	<input type="checkbox"/>	10. Attachment, Belonging	<input type="checkbox"/>
11. General Wellbeing	<input type="checkbox"/>	12. Safety, Security, and/or Risk	<input type="checkbox"/>
13. Other (please state):	<input type="checkbox"/>		<input type="checkbox"/>

Who would you recommend complete/fill out this outcome measure routinely?

1. Practitioners (specify)	<input type="checkbox"/>
2. Family/Whanau/Parents/ Caregivers	<input type="checkbox"/>
3. Client	<input type="checkbox"/>
4. Teacher	<input type="checkbox"/>
5. Peers/ Siblings	<input type="checkbox"/>
6. Other (specify)	<input type="checkbox"/>

14. What might prevent some practitioners from using measures to monitor change and/or measure outcomes?

Rate the statements below (using a 1—5 scale) according to how significant each is for preventing practitioners from using measures.

1 = least important 2 3 = moderately important 4 5 = most important

- Measures are not relevant in their professional practice
- The measures available are not satisfactory for use with HCN funded clients
- Norms and cutoff scores are not relevant for use in Aotearoa / New Zealand
- The measures are culturally inappropriate for use in Aotearoa / New Zealand
- The measures are not user-friendly (e.g., too complicated, incomprehensible)
- Expensive to purchase and use
- The qualifications required to purchase and/or administer measures
- Measures may become obsolete (e.g., new editions or versions)
- The time involved in training and learning a new measure or tool or system
- A limited awareness of the range of measures available
- Lack of sensitivity to change with a client
- Specific characteristics of the measures: too broad, too narrow etc.
- Measures can be misleading, out of context
- No training available
- Intentional bias (client, person administering), subjectivity
- Unavailable
- Other factors.....

15. What might be reasons for practitioners to use outcome measures?

Rate the statements below (using a 1—5 scale) according to importance of why practitioners should use measures

1 = least important 2 3 = moderately important 4 5 = most important

- Ease/simplicity of administration/completion
- Specific characteristics: developmentally appropriate, positive, multiple informants, comprehensive, valid for use with particular client
- Psychometric quality – validity, reliability
- Assists in diagnosis – screening, identifies needs, supports clinical interview
- Inexpensive, free to use
- Promotes discussion between client and clinician
- Verification purposes/ Use as a check of decisions
- Widely used
- Tracks changes/outcome
- Provides accountability
- Other factors.....

16. If you had enough time and resources (e.g., costs covered, no barriers) what would your ideal measurement tool(s) be, and why?

17. Think about the children/ adolescents you have seen in HCN. What sort of domains would you like to use measures for (please tick)?

1. Physical Health	<input type="checkbox"/>	2. Mental Health	<input type="checkbox"/>
3. Development	<input type="checkbox"/>	4. Independence	<input type="checkbox"/>
5. Social Interaction	<input type="checkbox"/>	6. Communication/Speech	<input type="checkbox"/>
7. Educational (learning, achievement)	<input type="checkbox"/>	8. Cultural Identity	<input type="checkbox"/>
9. Family/Whanau Relationships	<input type="checkbox"/>	10. Attachment, Belonging	<input type="checkbox"/>
11. General Wellbeing	<input type="checkbox"/>	12. Safety, Security, and/or Risk	<input type="checkbox"/>
13. Other (please state):	<input type="checkbox"/>		<input type="checkbox"/>

.....

SECTION B: FEEDBACK (OPTIONAL)

We are very interested in receiving your feedback about the survey and the broader research programme. We highly value your opinions.

18. Would a recommended set of measures for this population be useful for you in providing services to HCN clients?

19. What are additional issues that should be considered when we are reviewing and choosing measures to recommend for use with HCN clients?

20. What other comments would you like to make about this survey and our broader research programme?

Your participation in this research is very much appreciated. Please check that you have answered all the questions you can and return this questionnaire in the pre-paid envelope provided. Feel free to contact Shane Harvey at 06 356 9099, extn 7171, with any further comments or questions you may have.

THANK YOU FOR YOUR HELP

APPENDIX B – STUDY TWO: HCN FOCUS GROUPS

- Information Sheet
- Expression of Interest Form
- Consent Form



Massey University

Possible outcome measures for individualised plans for children and youth with high and complex needs:

Plan Advisor and Practitioner focus groups.

INFORMATION SHEET

Introduction

You are invited to participate in a research project that forms part of a broader research programme entitled ***Possible Outcome Measures for Individualised Plans for Children and Young People with High and Complex Needs***. Your experience and expertise is highly valued. We would very much appreciate hearing your views and invite you to take part in this project.

This particular phase of our project involves collecting the ideas, views and insights of both Plan Advisors and front-line staff of High and Complex Needs Exceptions Funded clients (HCN Clients) within a Focus Group setting. We will be running a single Focus Group for the Plan Advisors and a number of local Focus Groups for front-line practitioners.

What is this project about?

The overall research project is focussed on the evaluation of outcomes for children and youth with high and complex needs who are in receipt of Exceptions Funding. At the present time there are a range of local and organisational practices in use for monitoring change and measuring outcomes with this group of young people, but there is little uniformity. This makes it difficult to evaluate the relative effectiveness of intervention strategies, to highlight and learn from what is working well, or to provide further support when interventions are not as useful as anticipated. As the first step in developing a framework for evaluating outcomes it is important to consider what the current practices are, their strengths and deficiencies, and the views of practitioners experienced with this client group on a range of alternative evaluation options. We are running Focus Groups in order to provide a more interactive forum through which we can hear the views of experienced practitioners on this range of issues.

We are an independent research team based at Massey University, Palmerston North, and the Psychology Centre in Hamilton. We are being funded by the High and Complex Needs Unit to conduct this research.

Am I eligible to take part?

You are eligible to take part in this study if you are (i) a current Plan Advisor for the HCN Unit, or (ii) have previously been, or are currently part of an intersectoral team who is working with a High and Complex Needs Exceptions Funded client.

If you are a member of one of these grouping and would like to participate in this study please complete the *Expression of Interest Form* and mail/fax it back to us. One of the research team will contact you directly to answer any questions you may have, and provide details of the date/venue of the focus group meeting in your area.

What am I being asked to do?

If you choose to participate in this study, you will be asked to take part in an informal, semi-structured discussion with a small number of other Plan Advisors or practitioners. The topic of discussion, facilitated by the moderator, will relate to current and possible future practices associated with the evaluation of intervention outcomes. The discussion will take place in suitable professional premises in a convenient geographical location. Refreshments will be provided. It is anticipated that the focus group will run for 1½ - 2 hours.

We don't anticipate that there would be any risk or discomfort to you as a result of participation.

What will happen to my information?

The focus group discussion will be digitally audio recorded and transcribed. A second researcher will be present to assist the moderator and take notes regarding any features of the discussion that may not be captured by the recording. Participants will agree to preserving the confidentiality of opinions expressed by other members of the group, as well as respecting all views and perspectives.

All identifying material, such as signed consent forms, will be kept (separately) in a locked filing cabinet in the Psychology Clinic at Massey University. The transcripts will be stored in a locked filing cabinet and/or password protected electronic file at The Psychology Centre, Hamilton. They are being stored at a separate location simply because this is where the transcripts will be prepared from the audio files. The digital audio recordings will also be held as password protected electronic files at The Psychology Centre, and will be securely erased at the end of the research project. The transcripts will be held for five years following the conclusion of the research project, as is normal practice for research data, and then destroyed.

Information gathered from this survey will inform our recommendations to the HCN Intersectoral Unit on assessment approaches and measures for monitoring outcomes with High and Complex Needs Exceptions Funded clients. All participants are assured of anonymity in any written reports, etc.

What can I expect from the researchers?

If you decide to take part in this project, the researchers will respect your right to:

- ask any questions of the researchers about the study at any time during participation;
- decline to answer any particular question in the focus group;
- withdraw from the study at any time;
- provide information on the understanding that it is completely confidential to the researchers;
- ask for the audio recorder to be turned off at any time during the discussion or leave the focus group at any point;
- request removal or amendment of any part/s of the recording of the focus group;
- be given a summary of the findings.

Who can I speak with about taking part in this project?

If you have further questions or concerns, please do not hesitate to contact any of the following:

Ian Evans, Shane Harvey, Amber Barry, Simon Bennett
 School of Psychology
 Massey University
 Private Bag 11 222
 Palmerston North.
 Telephone 06 350 9099 xtn 2070

John Fitzgerald
 The Psychology Centre
 1st Floor, 2 Von Tempsky Street
 P.O. Box 4423,
 Hamilton 3247
 Telephone 07 834 1520

This project has been reviewed and approved by the Massey University Human Ethics Committee, PN Application 06/45. If you have any concerns about the conduct of this research, please contact Professor Sylvia V Rumball, Chair, Massey University Campus Human Ethics Committee: Palmerston North, telephone 06-350-5249, email humanethicspn@massey.ac.nz.



Massey University

Possible outcome measures for individualised plans for children and youth with high and complex needs: Plan Advisor and practitioner focus groups.

EXPRESSION OF INTEREST FORM

Having read the Information Sheet for the above project I am interested in participating in one of the focus groups. Please contact me with further information about the group running in my local area.

Name: _____

Contact Address: _____

Daytime telephone: _____

Email: _____

Employing agency (please circle): Health / Welfare / Education

Please post this form to: Ms Amber Barry
 School of Psychology
 Massey University
 Private Bag 11 222
 Palmerston North.

Or fax to Ms Amber Barry at 06 350 5673



Outcome measures for individualised plans for children and youth with high and complex needs: Plan Advisor and practitioner focus groups.

Participant Consent Form

This participant consent form will be held for a period of five (5) years

I have read the Information Sheet and have had the details of the study explained to me. My questions have been answered to my satisfaction, and I understand that I may ask further questions at any time.

I agree to the interview being audio recorded.

I agree to not disclose anything discussed in the Focus Group.

I agree to participate in this study under the conditions set out in the Information Sheet.

Signature: **Date:**

Full Name-printed:

APPENDIX C – STUDY FOUR: EXPERT AND PRACTITIONER CONSULTATION

- NCFAS
- Ohio Scales – P form
- Ohio Scales – W form
- Ohio Scales – Y form
- PSC – Y form
- SDQ – Parent of 4-10 year olds, Follow Up version
- SDQ – Teacher of 11-17 year olds, Follow Up version
- SDQ – Self-report for 11-17 year olds, Follow Up version
- YOQ – Parent form
- YOQ – Self-report form

NCFAS
North Carolina Family Assessment Scale
for Intensive Family Preservation Services (IFPS) Programs
Version 2.0

IFPS System ID#:	Date Intake Assessment Completed: / /
IFPS Worker:	Date Case Closure Assessment Completed: / /
Other Agency ID#:	Family Name:

Introduction

Each of the following scales is used to determine how a family is functioning. They also may be important to the level of imminent risk of out-of-home placement for this family in the context of family strengths and problems.

For each scale, rate its influence as a strength or problem for the family along a 6-point continuum, using the following schema: +2 = Clear Strength, +1 = Mild Strength, 0 = Baseline/Adequate, -1 = Mild Problem, -2 = Moderate Problem, and -3 = Serious Problem. To rate each scale, circle the appropriate number. "I" represents the rating given at intake, and "C" represents the rating at service or case closure.

The "overall" ratings (the ones in the shaded areas) should indicate your overall, composite rating in each of the five domains. The subscales represent areas of interest relating to the domain under which they appear (e.g., Housing Stability appears under domain A. Environment). The reliability and validity study of the NCFAS revealed that it is essential to rate each of the subscales before rating the overall domain scale. Use the definitions in the Definitions Manual to the NCFAS (Version 2.0 or higher) to help make your ratings.

Complete each of the ratings within 1 to 2 weeks of intake (I) and again within 1 to 2 weeks of service or case closure (C).

Many questions and issues of concern to practitioners are addressed in the User's Guide to the NCFAS (Version 2.0). Please also see the User's Guide for a discussion of the development and use of the Scale. The psychometric properties (reliability and validity) of the scale are also discussed in the User's Guide.

NCFAS: North Carolina Family Assessment Scale, Version 2.0. Kirk, R. S., and Reed Ashcraft, K, 06/98. This instrument is derived from previous versions based on the Family Assessment Form, developed at the Children's Bureau of Southern California, Michigan's Family Assessment of Needs Form, and four assessment instruments developed in North Carolina by Haven House (Raleigh), Home Remedies (Morganton), Methodist Home for Children (Raleigh), and the state Division of Mental Health, Developmental Disabilities and Substance Abuse Services. Special acknowledgments are due to Sandy Sladen and Judith Nelson at the Children's Bureau of Southern California and to researchers Jacquelyn McCroskey and William Meezan at U. of Southern California. Special thanks also are due to numerous local IFPS providers in North Carolina for participating in the ongoing development and field testing of the NCFAS. Domain specifications for the original NCFAS were based on the work of Meezan and McCroskey. Domains and subscales for Version 2.0 are based upon reliability and validity testing completed in the Fall of 1997.

(*) Re: asterisked items, theoretical and empirical support exists in the literature for the Parental Capabilities domain and the associated subscales, and several other subscales that either were not supported or examined independently in the 1997 reliability and validity study of the NCFAS. These items will be tested during future studies. See User's Guide to the NCFAS, Version 2.0, for additional information on scale construction and psychometrics.

A. Environment

	Clear Strength	Mild Strength	Baseline/Adequate	Mild Problem	Moderate Problem	Serious Problem
1. Overall Environment						
Intake (I)	+2	+1	0	-1	-2	-3
Closure (C)	+2	+1	0	-1	-2	-3
2. Housing Stability						
Intake (I)	+2	+1	0	-1	-2	-3
Closure (C)	+2	+1	0	-1	-2	-3
3. Safety in the Community						
Intake (I)	+2	+1	0	-1	-2	-3
Closure (C)	+2	+1	0	-1	-2	-3
4. Habitability of Housing						
Intake (I)	+2	+1	0	-1	-2	-3
Closure (C)	+2	+1	0	-1	-2	-3
5. Income/Employment						
Intake (I)	+2	+1	0	-1	-2	-3
Closure (C)	+2	+1	0	-1	-2	-3
6. Financial Management						
Intake (I)	+2	+1	0	-1	-2	-3
Closure (C)	+2	+1	0	-1	-2	-3
7. Food and Nutrition						
Intake (I)	+2	+1	0	-1	-2	-3
Closure (C)	+2	+1	0	-1	-2	-3
8. Personal Hygiene						
Intake (I)	+2	+1	0	-1	-2	-3
Closure (C)	+2	+1	0	-1	-2	-3
9. Transportation						
Intake (I)	+2	+1	0	-1	-2	-3
Closure (C)	+2	+1	0	-1	-2	-3
10. Learning Environment						
Intake (I)	+2	+1	0	-1	-2	-3
Closure (C)	+2	+1	0	-1	-2	-3

B. Parental Capabilities

Note: This section refers to biological parent(s), if present, or current caregiver(s).

	Clear Strength	Mild Strength	Baseline/Adequate	Mild Problem	Moderate Problem	Serious Problem
1. Overall Parental Capabilities						
Intake (I)	+2	+1	0	-1	-2	-3
Closure (C)	+2	+1	0	-1	-2	-3
2. Supervision of Child(ren)						
Intake (I)	+2	+1	0	-1	-2	-3
Closure (C)	+2	+1	0	-1	-2	-3
3. Disciplinary Practices						
Intake (I)	+2	+1	0	-1	-2	-3
Closure (C)	+2	+1	0	-1	-2	-3
4. Provision of Developmental/Enrichment Opportunities						
Intake (I)	+2	+1	0	-1	-2	-3
Closure (C)	+2	+1	0	-1	-2	-3
5. Parent(s)/Caregiver(s') Mental Health						
Intake (I)	+2	+1	0	-1	-2	-3
Closure (C)	+2	+1	0	-1	-2	-3
6. Parent(s)/Caregiver(s') Physical Health						
Intake (I)	+2	+1	0	-1	-2	-3
Closure (C)	+2	+1	0	-1	-2	-3
7. Parent(s)/Caregiver(s') Use of Drugs/Alcohol						
Intake (I)	+2	+1	0	-1	-2	-3
Closure (C)	+2	+1	0	-1	-2	-3

C. Family Interactions

Note: This section refers to family members living in the same or different households.

		Clear Strength	Mild Strength	Baseline/ Adequate	Mild Problem	Moderate Problem	Serious Problem
1. Overall Family Interactions							
Intake (I)		+2	+1	0	-1	-2	-3
Closure (C)		+2	+1	0	-1	-2	-3
2. Bonding with Child(ren)							
Intake (I)		+2	+1	0	-1	-2	-3
Closure (C)		+2	+1	0	-1	-2	-3
3. Expectations of the Child(ren)							
Intake (I)		+2	+1	0	-1	-2	-3
Closure (C)		+2	+1	0	-1	-2	-3
4. Mutual Support Within the Family							
Intake (I)		+2	+1	0	-1	-2	-3
Closure (C)		+2	+1	0	-1	-2	-3
5. Relationship Between Parents/Caregivers*							
Intake (I)	N/A	+2	+1	0	-1	-2	-3
Closure (C)	N/A	+2	+1	0	-1	-2	-3

D. Family Safety

Note: This section refers to family members living in the same or different households

		Clear Strength	Mild Strength	Baseline/ Adequate	Mild Problem	Moderate Problem	Serious Problem	
1. Overall Family Safety								
	Intake (I)	+2	+1	0	-1	-2	-3	
	Closure (C)	+2	+1	0	-1	-2	-3	
2. Absence/Presence of Physical Abuse of Child(ren)*								
	Intake (I)	N/A	+2	+1	0	-1	-2	-3
	Closure (C)	N/A	+2	+1	0	-1	-2	-3
3. Absence/Presence of Sexual Abuse of Child(ren)								
	Intake (I)	+2	+1	0	-1	-2	-3	
	Closure (C)	+2	+1	0	-1	-2	-3	
4. Absence/Presence of Emotional Abuse of Child(ren)								
	Intake (I)	+2	+1	0	-1	-2	-3	
	Closure (C)	+2	+1	0	-1	-2	-3	
5. Absence/Presence of Neglect of Child(ren)								
	Intake (I)	+2	+1	0	-1	-2	-3	
	Closure (C)	+2	+1	0	-1	-2	-3	
6. Absence/Presence of Domestic Violence Between Parents/Caregivers								
	Intake (I)	+2	+1	0	-1	-2	-3	
	Closure (C)	+2	+1	0	-1	-2	-3	

E. Child Well-Being

Note: This section pertains to the imminent risk child(ren).

		Clear Strength	Mild Strength	Baseline/ Adequate	Mild Problem	Moderate Problem	Serious Problem
1. Overall Child Well-Being							
Intake (I)		+2	+1	0	-1	-2	-3
Closure (C)		+2	+1	0	-1	-2	-3
2. Child(ren)'s Mental Health							
Intake (I)		+2	+1	0	-1	-2	-3
Closure (C)		+2	+1	0	-1	-2	-3
3. Child(ren)'s Behavior							
Intake (I)		+2	+1	0	-1	-2	-3
Closure (C)		+2	+1	0	-1	-2	-3
4. School Performance							
Intake (I)		+2	+1	0	-1	-2	-3
Closure (C)		+2	+1	0	-1	-2	-3
5. Relationship with Parent(s)/Caregiver(s)							
Intake (I)		+2	+1	0	-1	-2	-3
Closure (C)		+2	+1	0	-1	-2	-3
6. Relationship with Sibling(s)							
Intake (I)	N/A	+2	+1	0	-1	-2	-3
Closure (C)	N/A	+2	+1	0	-1	-2	-3
7. Relationship with Peers							
Intake (I)	N/A	+2	+1	0	-1	-2	-3
Closure (C)	N/A	+2	+1	0	-1	-2	-3
8. Cooperation/Motivation To Maintain the Family							
Intake (I)		+2	+1	0	-1	-2	-3
Closure (C)		+2	+1	0	-1	-2	-3



Ohio Mental Health Consumer Outcomes System
Ohio Youth Problem, Functioning, and Satisfaction Scales
 Parent Rating – Short Form

P

Child's Name: _____ Date: _____ Child's Grade: _____ ID#: _____
Completed by Agency

Child's Date of Birth: _____ Child's Sex: Male Female Child's Race: _____

Form Completed By: Mother Father Step-mother Step-father Other: _____

Instructions: Please rate the degree to which your child has experienced the following problems in the past 30 days.	Not at All	Once or Twice	Several Times	Often	Most of the Time	All of the Time
1. Arguing with others	0	1	2	3	4	5
2. Getting into fights	0	1	2	3	4	5
3. Yelling, swearing, or screaming at others	0	1	2	3	4	5
4. Fits of anger	0	1	2	3	4	5
5. Refusing to do things teachers or parents ask	0	1	2	3	4	5
6. Causing trouble for no reason	0	1	2	3	4	5
7. Using drugs or alcohol	0	1	2	3	4	5
8. Breaking rules or breaking the law (out past curfew, stealing)	0	1	2	3	4	5
9. Skipping school or classes	0	1	2	3	4	5
10. Lying	0	1	2	3	4	5
11. Can't seem to sit still, having too much energy	0	1	2	3	4	5
12. Hurting self (cutting or scratching self, taking pills)	0	1	2	3	4	5
13. Talking or thinking about death	0	1	2	3	4	5
14. Feeling worthless or useless	0	1	2	3	4	5
15. Feeling lonely and having no friends	0	1	2	3	4	5
16. Feeling anxious or fearful	0	1	2	3	4	5
17. Worrying that something bad is going to happen	0	1	2	3	4	5
18. Feeling sad or depressed	0	1	2	3	4	5
19. Nightmares	0	1	2	3	4	5
20. Eating problems	0	1	2	3	4	5

(Add ratings together) Total _____

<p>Instructions: Please circle your response to each question.</p> <ol style="list-style-type: none"> 1. Overall, how satisfied are you with your relationship with your child right now? <ol style="list-style-type: none"> 1. Extremely satisfied 2. Moderately satisfied 3. Somewhat satisfied 4. Somewhat dissatisfied 5. Moderately dissatisfied 6. Extremely dissatisfied 2. How capable of dealing with your child's problems do you feel right now? <ol style="list-style-type: none"> 1. Extremely capable 2. Moderately capable 3. Somewhat capable 4. Somewhat incapable 5. Moderately incapable 6. Extremely incapable 3. How much stress or pressure is in your life right now? <ol style="list-style-type: none"> 1. Very little 2. Some 3. Quite a bit 4. A moderate amount 5. A great deal 6. Unbearable amounts 4. How optimistic are you about your child's future right now? <ol style="list-style-type: none"> 1. The future looks very bright 2. The future looks somewhat bright 3. The future looks OK 4. The future looks both good and bad 5. The future looks bad 6. The future looks very bad <p style="text-align: right;">Total: _____</p>	<p>Instructions: Please circle your response to each question.</p> <ol style="list-style-type: none"> 1. How satisfied are you with the mental health services your child has received so far? <ol style="list-style-type: none"> 1. Extremely satisfied 2. Moderately satisfied 3. Somewhat satisfied 4. Somewhat dissatisfied 5. Moderately dissatisfied 6. Extremely dissatisfied 2. To what degree have you been included in the treatment planning process for your child? <ol style="list-style-type: none"> 1. A great deal 2. Moderately 3. Quite a bit 4. Somewhat 5. A little 6. Not at all 3. Mental health workers involved in my case listen to and value my ideas about treatment planning for my child. <ol style="list-style-type: none"> 1. A great deal 2. Moderately 3. Quite a bit 4. Somewhat 5. A little 6. Not at all 4. To what extent does your child's treatment plan include your ideas about your child's treatment needs? <ol style="list-style-type: none"> 1. A great deal 2. Moderately 3. Quite a bit 4. Somewhat 5. A little 6. Not at all <p style="text-align: right;">Total: _____</p>
--	--

Instructions: Please rate the degree to which your child's problems affect his or her current ability in everyday activities. Consider your child's current level of functioning.	Extreme Troubles	Quite a Few Troubles	Some Troubles	OK	Doing Very Well
1. Getting along with friends	0	1	2	3	4
2. Getting along with family	0	1	2	3	4
3. Dating or developing relationships with boyfriends or girlfriends	0	1	2	3	4
4. Getting along with adults outside the family (teachers, principal)	0	1	2	3	4
5. Keeping neat and clean, looking good	0	1	2	3	4
6. Caring for health needs and keeping good health habits (taking medicines or brushing teeth)	0	1	2	3	4
7. Controlling emotions and staying out of trouble	0	1	2	3	4
8. Being motivated and finishing projects	0	1	2	3	4
9. Participating in hobbies (baseball cards, coins, stamps, art)	0	1	2	3	4
10. Participating in recreational activities (sports, swimming, bike riding)	0	1	2	3	4
11. Completing household chores (cleaning room, other chores)	0	1	2	3	4
12. Attending school and getting passing grades in school	0	1	2	3	4
13. Learning skills that will be useful for future jobs	0	1	2	3	4
14. Feeling good about self	0	1	2	3	4
15. Thinking clearly and making good decisions	0	1	2	3	4
16. Concentrating, paying attention, and completing tasks	0	1	2	3	4
17. Earning money and learning how to use money wisely	0	1	2	3	4
18. Doing things without supervision or restrictions	0	1	2	3	4
19. Accepting responsibility for actions	0	1	2	3	4
20. Ability to express feelings	0	1	2	3	4



Ohio Mental Health Consumer Outcomes System
Ohio Youth Problem, Functioning, and Satisfaction Scales
 Agency Worker Rating – Short Form

W

Child's Name: _____ Date: _____ Child's Grade: _____ ID#: _____

Child's Date of Birth: _____ Child's Sex: Male Female Child's Race: _____

Form Completed By: _____ Case Manager Therapist Other: _____

Instructions: Please rate the degree to which the designated child has experienced the following problems in the past 30 days.	Not at All	Once or Twice	Several Times	Often	Most of the Time	All of the Time
1. Arguing with others	0	1	2	3	4	5
2. Getting into fights	0	1	2	3	4	5
3. Yelling, swearing, or screaming at others	0	1	2	3	4	5
4. Fits of anger	0	1	2	3	4	5
5. Refusing to do things teachers or parents ask	0	1	2	3	4	5
6. Causing trouble for no reason	0	1	2	3	4	5
7. Using drugs or alcohol	0	1	2	3	4	5
8. Breaking rules or breaking the law (out past curfew, stealing)	0	1	2	3	4	5
9. Skipping school or classes	0	1	2	3	4	5
10. Lying	0	1	2	3	4	5
11. Can't seem to sit still, having too much energy	0	1	2	3	4	5
12. Hurting self (cutting or scratching self, taking pills)	0	1	2	3	4	5
13. Talking or thinking about death	0	1	2	3	4	5
14. Feeling worthless or useless	0	1	2	3	4	5
15. Feeling lonely and having no friends	0	1	2	3	4	5
16. Feeling anxious or fearful	0	1	2	3	4	5
17. Worrying that something bad is going to happen	0	1	2	3	4	5
18. Feeling sad or depressed	0	1	2	3	4	5
19. Nightmares	0	1	2	3	4	5
20. Eating problems	0	1	2	3	4	5

(Add ratings together) Total _____

ROLES: Enter the number of days the youth was placed in each of the following settings during the past 90 days. (For example, the youth may have been in a detention center for 3 days, a group home for 7 days and with the biological mother for 80 days.)

_____ Jail	_____ Foster Care
_____ Juvenile Detention Center	_____ Supervised Independent Living
_____ Inpatient Psychiatric Hospital	_____ Home of a Family Friend
_____ Drug/Alcohol Rehabilitation Center	_____ Adoptive Home
_____ Medical Hospital	_____ Home of a Relative
_____ Residential Treatment	_____ School Dormitory
_____ Group Emergency Shelter	_____ Biological Father
_____ Residential Job Corp/Vocational Center	_____ Biological Mother
_____ Group Home	_____ Two Biological Parents
_____ Therapeutic Foster Care	_____ Independent Living with Friend
_____ Individual Home Emergency Shelter	_____ Independent Living by Self
_____ Specialized Foster Care	

90 (Total for the two columns should equal 90)

Markers:	Number in Past 90 Days
School Placement: _____	Arrests _____
Current Psychoactive Medications: _____	Suspensions from school _____
_____	Days in Detention _____
	Days of School Missed _____
	Self-Harm Attempts _____

Instructions: Please circle the number corresponding to the designated youth's current level of functioning in each area.	Extreme Troubles	Quite a Few Troubles	Some Troubles	OK	Doing Very Well
1. Getting along with friends	0	1	2	3	4
2. Getting along with family	0	1	2	3	4
3. Dating or developing relationships with boyfriends or girlfriends	0	1	2	3	4
4. Getting along with adults outside the family (teachers, principal)	0	1	2	3	4
5. Keeping neat and clean, looking good	0	1	2	3	4
6. Caring for health needs and keeping good health habits (taking medicines or brushing teeth)	0	1	2	3	4
7. Controlling emotions and staying out of trouble	0	1	2	3	4
8. Being motivated and finishing projects	0	1	2	3	4
9. Participating in hobbies (baseball cards, coins, stamps, art)	0	1	2	3	4
10. Participating in recreational activities (sports, swimming, bike riding)	0	1	2	3	4
11. Completing household chores (cleaning room, other chores)	0	1	2	3	4
12. Attending school and getting passing grades in school	0	1	2	3	4
13. Learning skills that will be useful for future jobs	0	1	2	3	4
14. Feeling good about self	0	1	2	3	4
15. Thinking clearly and making good decisions	0	1	2	3	4
16. Concentrating, paying attention, and completing tasks	0	1	2	3	4
17. Earning money and learning how to use money wisely	0	1	2	3	4
18. Doing things without supervision or restrictions	0	1	2	3	4
19. Accepting responsibility for actions	0	1	2	3	4
20. Ability to express feelings	0	1	2	3	4

Copyright © Benjamin M. Ogles & Southern Consortium for Children January 2000 (Worker-2) (Add ratings together) Total _____



Ohio Mental Health Consumer Outcomes System
Ohio Youth Problem, Functioning, and Satisfaction Scales
 Youth Rating – Short Form (Ages 12-18)

Y

Name: _____ Date: _____ Grade: _____

ID#: _____ Completed by Agency _____

Date of Birth: _____ Sex: Male Female Race: _____

Instructions: Please rate the degree to which you have experienced the following problems in the past 30 days.	Not at All	Once or Twice	Several Times	Often	Most of the Time	All of the Time
1. Arguing with others	0	1	2	3	4	5
2. Getting into fights	0	1	2	3	4	5
3. Yelling, swearing, or screaming at others	0	1	2	3	4	5
4. Fits of anger	0	1	2	3	4	5
5. Refusing to do things teachers or parents ask	0	1	2	3	4	5
6. Causing trouble for no reason	0	1	2	3	4	5
7. Using drugs or alcohol	0	1	2	3	4	5
8. Breaking rules or breaking the law (out past curfew, stealing)	0	1	2	3	4	5
9. Skipping school or classes	0	1	2	3	4	5
10. Lying	0	1	2	3	4	5
11. Can't seem to sit still, having too much energy	0	1	2	3	4	5
12. Hurting self (cutting or scratching self, taking pills)	0	1	2	3	4	5
13. Talking or thinking about death	0	1	2	3	4	5
14. Feeling worthless or useless	0	1	2	3	4	5
15. Feeling lonely and having no friends	0	1	2	3	4	5
16. Feeling anxious or fearful	0	1	2	3	4	5
17. Worrying that something bad is going to happen	0	1	2	3	4	5
18. Feeling sad or depressed	0	1	2	3	4	5
19. Nightmares	0	1	2	3	4	5
20. Eating problems	0	1	2	3	4	5

(Add ratings together) Total _____

<p>Instructions: Please circle your response to each question.</p> <ol style="list-style-type: none"> 1. Overall, how satisfied are you with your life right now? <ol style="list-style-type: none"> 1. Extremely satisfied 2. Moderately satisfied 3. Somewhat satisfied 4. Somewhat dissatisfied 5. Moderately dissatisfied 6. Extremely dissatisfied 2. How energetic and healthy do you feel right now? <ol style="list-style-type: none"> 1. Extremely healthy 2. Moderately healthy 3. Somewhat healthy 4. Somewhat unhealthy 5. Moderately unhealthy 6. Extremely unhealthy 3. How much stress or pressure is in your life right now? <ol style="list-style-type: none"> 1. Very little stress 2. Some stress 3. Quite a bit of stress 4. A moderate amount of stress 5. A great deal of stress 6. Unbearable amounts of stress 4. How optimistic are you about the future? <ol style="list-style-type: none"> 1. The future looks very bright 2. The future looks somewhat bright 3. The future looks OK 4. The future looks both good and bad 5. The future looks bad 6. The future looks very bad <p style="text-align: right;">Total: _____</p>	<p>Instructions: Please circle your response to each question.</p> <ol style="list-style-type: none"> 1. How satisfied are you with the mental health services you have received so far? <ol style="list-style-type: none"> 1. Extremely satisfied 2. Moderately satisfied 3. Somewhat satisfied 4. Somewhat dissatisfied 5. Moderately dissatisfied 6. Extremely dissatisfied 2. How much are you included in deciding your treatment? <ol style="list-style-type: none"> 1. A great deal 2. Moderately 3. Quite a bit 4. Somewhat 5. A little 6. Not at all 3. Mental health workers involved in my case listen to me and know what I want. <ol style="list-style-type: none"> 1. A great deal 2. Moderately 3. Quite a bit 4. Somewhat 5. A little 6. Not at all 4. I have a lot of say about what happens in my treatment. <ol style="list-style-type: none"> 1. A great deal 2. Moderately 3. Quite a bit 4. Somewhat 5. A little 6. Not at all <p style="text-align: right;">Total: _____</p>
---	--

Instructions: Below are some ways your problems might get in the way of your ability to do everyday activities. Read each item and circle the number that best describes your current situation.	Extreme Troubles	Quite a Few Troubles	Some Troubles	OK	Doing Very Well
1. Getting along with friends	0	1	2	3	4
2. Getting along with family	0	1	2	3	4
3. Dating or developing relationships with boyfriends or girlfriends	0	1	2	3	4
4. Getting along with adults outside the family (teachers, principal)	0	1	2	3	4
5. Keeping neat and clean, looking good	0	1	2	3	4
6. Caring for health needs and keeping good health habits (taking medicines or brushing teeth)	0	1	2	3	4
7. Controlling emotions and staying out of trouble	0	1	2	3	4
8. Being motivated and finishing projects	0	1	2	3	4
9. Participating in hobbies (baseball cards, coins, stamps, art)	0	1	2	3	4
10. Participating in recreational activities (sports, swimming, bike riding)	0	1	2	3	4
11. Completing household chores (cleaning room, other chores)	0	1	2	3	4
12. Attending school and getting passing grades in school	0	1	2	3	4
13. Learning skills that will be useful for future jobs	0	1	2	3	4
14. Feeling good about self	0	1	2	3	4
15. Thinking clearly and making good decisions	0	1	2	3	4
16. Concentrating, paying attention, and completing tasks	0	1	2	3	4
17. Earning money and learning how to use money wisely	0	1	2	3	4
18. Doing things without supervision or restrictions	0	1	2	3	4
19. Accepting responsibility for actions	0	1	2	3	4
20. Ability to express feelings	0	1	2	3	4

(Add ratings together) Total _____

Pediatric Symptom Checklist - Youth Report (Y-PSC)

Please mark under the heading that best fits you:

	Never	Sometimes	Often
1. Complain of aches or pains.....	—	—	—
2. Spend more time alone.....	—	—	—
3. Tire easily, little energy.....	—	—	—
4. Fidgety, unable to sit still.....	—	—	—
5. Have trouble with teacher.....	—	—	—
6. Less interested in school.....	—	—	—
7. Act as if driven by motor.....	—	—	—
8. Daydream too much.....	—	—	—
9. Distract easily.....	—	—	—
10. Are afraid of new situations.....	—	—	—
11. Feel sad, unhappy.....	—	—	—
12. Are irritable, angry.....	—	—	—
13. Feel hopeless.....	—	—	—
14. Have trouble concentrating.....	—	—	—
15. Less interested in friends.....	—	—	—
16. Fight with other children.....	—	—	—
17. Absent from school.....	—	—	—
18. School grades dropping.....	—	—	—
19. Down on yourself.....	—	—	—
20. Visit doctor with doctor finding nothing wrong.....	—	—	—
21. Have trouble sleeping.....	—	—	—
22. Worry a lot.....	—	—	—
23. Want to be with parent more than before.....	—	—	—
24. Feel that you are bad.....	—	—	—
25. Take unnecessary risks.....	—	—	—
26. Get hurt frequently.....	—	—	—
27. Seem to be having less fun.....	—	—	—
28. Act younger than children your age.....	—	—	—
29. Do not listen to rules.....	—	—	—
30. Do not show feelings.....	—	—	—
31. Do not understand other people's feelings.....	—	—	—
32. Tease others.....	—	—	—
33. Blame others for your troubles.....	—	—	—
34. Take things that do not belong to you.....	—	—	—
35. Refuse to share.....	—	—	—

Strengths and Difficulties Questionnaire

P4-10
FOLLOW-UP

For each item, please mark the box for Not True, Somewhat True or Certainly True. It would help us if you answered all items as best you can even if you are not absolutely certain. Please give your answers on the basis of your child's behaviour over the last month.

Your child's name

Male/Female

Date of birth

	Not True	Somewhat True	Certainly True
Considerate of other people's feelings	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Restless, overactive, cannot stay still for long	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Often complains of headaches, stomach-aches or sickness	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Shares readily with other children, for example toys, treats, pencils	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Often loses temper	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Rather solitary, prefers to play alone	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Generally well behaved, usually does what adults request	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Many worries or often seems worried	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Helpful if someone is hurt, upset or feeling ill	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Constantly fidgeting or squirming	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Has at least one good friend	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Often fights with other children or bullies them	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Often unhappy, depressed or tearful	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Generally liked by other children	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Easily distracted, concentration wanders	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Nervous or clingy in new situations, easily loses confidence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kind to younger children	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Often lies or cheats	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Picked on or bullied by other children	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Often volunteers to help others (parents, teachers, other children)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Thinks things out before acting	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Steals from home, school or elsewhere	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Gets along better with adults than with other children	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Many fears, easily scared	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Good attention span, sees chores or homework through to the end	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Do you have any other comments or concerns?

Please turn over - there are a few more questions on the other side

Since coming to the service, are your child's problems:

Much worse	A bit worse	About the same	A bit better	Much better
<input type="checkbox"/>				

Has coming to the service been helpful in other ways, e.g. providing information or making the problems more bearable?

Not at all	A little	A medium amount	A great deal
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Overall, do you think that your child has difficulties in any of the following areas: emotions, concentration, behaviour or being able to get along with other people?

No	Yes - minor difficulties	Yes - definite difficulties	Yes - severe difficulties
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

If you have answered "Yes", please answer the following questions about these difficulties:

- Do the difficulties upset or distress your child?

Not at all	A little	A medium amount	A great deal
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

- Do the difficulties interfere with your child's everyday life in the following areas?

	Not at all	A little	A medium amount	A great deal
HOME LIFE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
FRIENDSHIPS	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
CLASSROOM LEARNING	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
LEISURE ACTIVITIES	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

- Do the difficulties put a burden on you or the family as a whole?

Not at all	A little	A medium amount	A great deal
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Signature Date

Mother / Father / Other (please specify:)

Thank you very much for your help

Strengths and Difficulties Questionnaire

T11-17
FOLLOW-UP

For each item, please mark the box for Not True, Somewhat True or Certainly True. It would help us if you answered all items as best you can even if you are not absolutely certain. Please give your answers on the basis of this student's behaviour over the last month.

Student's name

Male/Female

Date of birth

	Not True	Somewhat True	Certainly True
Considerate of other people's feelings	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Restless, overactive, cannot stay still for long	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Often complains of headaches, stomach-aches or sickness	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Shares readily with other young people, for example pencils, books, food	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Often loses temper	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Would rather be alone than with other young people	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Generally well behaved, usually does what adults request	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Many worries or often seems worried	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Helpful if someone is hurt, upset or feeling ill	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Constantly fidgeting or squirming	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Has at least one good friend	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Often fights with other young people or bullies them	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Often unhappy, depressed or tearful	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Generally liked by other young people	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Easily distracted, concentration wanders	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Nervous in new situations, easily loses confidence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kind to younger children	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Often lies or cheats	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Picked on or bullied by other young people	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Often volunteers to help others (parents, teachers, children)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Thinks things out before acting	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Steals from home, school or elsewhere	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Gets along better with adults than with other young people	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Many fears, easily scared	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Good attention span, sees tasks through to the end	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Do you have any other comments or concerns?

Please turn over - there are a few more questions on the other side

Since coming to the service, are this student's problems:

Much worse	A bit worse	About the same	A bit better	Much better
<input type="checkbox"/>				

Has coming to the service been helpful in other ways, e.g. providing information or making the problems more bearable?

Not at all	A little	A medium amount	A great deal
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Overall, do you think that this student has difficulties in any of the following areas: emotions, concentration, behaviour or being able to get along with other people?

No	Yes - minor difficulties	Yes - definite difficulties	Yes - severe difficulties
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

If you have answered "Yes", please answer the following questions about these difficulties:

- Do the difficulties upset or distress this student?

Not at all	A little	A medium amount	A great deal
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

- Do the difficulties interfere with this student's everyday life in the following areas?

	Not at all	A little	A medium amount	A great deal
PEER RELATIONSHIPS	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
CLASSROOM LEARNING	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

- Do the difficulties put a burden on you or the class as a whole?

Not at all	A little	A medium amount	A great deal
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Signature Date

Thank you very much for your help

Strengths and Difficulties Questionnaire

S11-17
FOLLOW-UP

For each item, please mark the box for Not True, Somewhat True or Certainly True. It would help us if you answered all items as best you can even if you are not absolutely certain. Please give your answers on the basis of how things have been for you over the last month.

Your name

Male/Female

Date of birth

	Not True	Somewhat True	Certainly True
I try to be nice to other people. I care about their feelings	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I am restless, I cannot stay still for long	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I get a lot of headaches, stomach-aches or sickness	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I usually share with others, for example CDs, games, food	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I get very angry and often lose my temper	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I would rather be alone than with people of my age	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I usually do as I am told	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I worry a lot	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I am helpful if someone is hurt, upset or feeling ill	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I am constantly fidgeting or squirming	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I have one good friend or more	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I fight a lot. I can make other people do what I want	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I am often unhappy, depressed or tearful	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other people my age generally like me	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I am easily distracted, I find it difficult to concentrate	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I am nervous in new situations. I easily lose confidence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I am kind to younger children	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I am often accused of lying or cheating	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other children or young people pick on me or bully me	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I often volunteer to help others (parents, teachers, children)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I think before I do things	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I take things that are not mine from home, school or elsewhere	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I get along better with adults than with people my own age	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I have many fears, I am easily scared	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I finish the work I'm doing. My attention is good	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Do you have any other comments or concerns?

Please turn over - there are a few more questions on the other side

Since coming to the service, are your problems:

Much worse	A bit worse	About the same	A bit better	Much better
<input type="checkbox"/>				

Has coming to the service been helpful in other ways, e.g. providing information or making the problems more bearable?

Not at all	A little	A medium amount	A great deal
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Overall, do you think that you have difficulties in any of the following areas: emotions, concentration, behaviour or being able to get along with other people?

No	Yes - minor difficulties	Yes - definite difficulties	Yes - severe difficulties
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

If you have answered "Yes", please answer the following questions about these difficulties:

- Do the difficulties upset or distress you?

Not at all	A little	A medium amount	A great deal
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

- Do the difficulties interfere with your everyday life in the following areas?

	Not at all	A little	A medium amount	A great deal
HOME LIFE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
FRIENDSHIPS	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
CLASSROOM LEARNING	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
LEISURE ACTIVITIES	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

- Do the difficulties make it harder for those around you (family, friends, teachers, etc.)?

Not at all	A little	A medium amount	A great deal
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Your Signature

Today's Date

Thank you very much for your help

Youth Outcome Questionnaire (Y-OQ® 2.01)

Name _____ Date _____

PURPOSE: The Y-OQ® 2.01 is designed to describe a wide range of troublesome situations, behaviors, and moods that are common in children and adolescents. You may discover that some of the items do not apply to your child's current situation. If so, please do not leave these items blank but check the "Never or almost never" category. When you begin to complete the Y-OQ® 2.01 you will see that you can easily make your child look as healthy or unhealthy as you wish. Please do not do that. If you are as accurate as possible it is more likely that you will be able to receive the help that you are seeking for your child.

DIRECTIONS:

- Read each statement carefully
- Decide how true this statement is for your child during the past 7 days.
- Completely fill the circle that most accurately describes your child during the past week
- Fill in only one answer for each statement and erase unwanted marks clearly.

Developed by Gary M. Burlingame, Ph.D., Gawan Wells, Ph.D. and Michael J. Lambert, Ph.D.

© Copyright 1996 American Professional Credentialing Services LLC. All Rights Reserved. License Required For All Uses

For More Information Contact:
 AMERICAN PROFESSIONAL CREDENTIALING SERVICES LLC
 PO Box 970354
 Orem, Utah 84097-0354

E-MAIL: APCSA@OQFAMILY.COM
 WEB: WWW.OQFAMILY.COM
 TOLL-FREE: 1-888-MHI SCORE (1-888-647-2673)
 FAX: 1-801-434-9730

	Never or Almost Never	Rarely	Sometimes	Frequently	Almost Always or Always
1. My child wants to be alone more than other children of the same age	○	○	○	○	○
2. My child complains of dizziness or headaches	○	○	○	○	○
3. My child doesn't participate in activities that were previously enjoyable	○	○	○	○	○
4. My child argues or is verbally disrespectful	○	○	○	○	○
5. My child is more fearful than other children of the same age	○	○	○	○	○
6. My child cuts school or is truant	○	○	○	○	○
7. My child cooperates with rules and expectations	○	○	○	○	○
8. My child has difficulty completing assignments, or completes them carelessly	○	○	○	○	○
9. My child complains or whines about things being unfair	○	○	○	○	○
10. My child experiences trouble with her/his bowels, such as constipation or diarrhea	○	○	○	○	○
11. My child gets into physical fights with peers or family members	○	○	○	○	○
12. My child worries and can't get certain ideas off his/her mind	○	○	○	○	○
13. My child steals or lies	○	○	○	○	○
14. My child is fidgety, restless, or hyperactive	○	○	○	○	○
15. My child seems anxious or nervous	○	○	○	○	○
16. My child communicates in a pleasant and appropriate manner	○	○	○	○	○
17. My child seems tense, easily startled	○	○	○	○	○
18. My child soils or wets self	○	○	○	○	○
19. My child is aggressive toward adults	○	○	○	○	○
20. My child sees, hears, or believes things that are not real	○	○	○	○	○
21. My child has participated in self-harm (e.g. cutting or scratching self, attempting suicide)	○	○	○	○	○
22. My child uses alcohol or drugs	○	○	○	○	○
23. My child seems unable to get organized	○	○	○	○	○
24. My child enjoys relationships with family and friends	○	○	○	○	○
25. My child appears sad or unhappy	○	○	○	○	○
26. My child experiences pain or weakness in muscles or joints	○	○	○	○	○
27. My child has a negative, distrustful attitude toward friends, family members, or other adults	○	○	○	○	○
28. My child believes that others are trying to hurt him/her even when they are not	○	○	○	○	○
29. My child threatens to, or has run away from home	○	○	○	○	○
30. My child experiences rapidly changing and strong emotions	○	○	○	○	○

Youth Outcome Questionnaire (Y-OQ® 2.01)

Name _____	Date _____	Never or Almost Never	Rarely	Sometimes	Frequently	Almost Always or Always
<p>PURPOSE: The Y-OQ® 2.01 is designed to describe a wide range of troublesome situations, behaviors, and moods that are common in children and adolescents. You may discover that some of the items do not apply to your child's current situation. If so, <u>please do not leave these items blank</u> but check the "Never or almost never" category. When you begin to complete the Y-OQ® 2.01 you will see that you can easily make your child look as healthy or unhealthy as you wish. <u>Please do not do that.</u> If you are as accurate as possible it is more likely that you will be able to receive the help that you are seeking for your child.</p> <p>DIRECTIONS:</p> <ul style="list-style-type: none"> ▪ Read each statement carefully ▪ Decide how true this statement is for your child during the past 7 days. ▪ Completely fill the circle that most accurately describes your child during the past week. ▪ Check only one answer for each statement and erase unwanted marks clearly. <div style="border: 1px solid black; padding: 5px; margin-top: 10px;"> <p>Developed by Gary M. Burlingame, Ph.D., Gawain Wells, Ph.D. and Michael J. Lambert, Ph.D.</p> <p>© Copyright 1996 American Professional Credentialing Services LLC. All Rights Reserved. License Required For All Uses</p> <p>For More Information Contact: AMERICAN PROFESSIONAL CREDENTIALING SERVICES LLC PO Box 970354 Orem, Utah 84097-0354</p> <p>E-MAIL: APCS@OQFAMILY.COM</p> <p>WEB: WWW.OQFAMILY.COM TOLL-FREE: 1-888-MHI SCORE, (1-888-647-2673) FAX: 1-801-434-9730</p> </div>	31. My child deliberately breaks rules, laws, or expectations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	32. My child appears happy with her/himself	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	33. My child sulks, pouts, or cries more than other children of the same age.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	34. My child pulls away from family or friends	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	35. My child complains of stomach pain or feeling sick more..... than other children of the same age.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	36. My child doesn't have or keep friends	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	37. My child has friends of whom I don't approve	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	38. My child believes that others can hear her/his thoughts..... or that s/he can hear the thoughts of others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	39. My child engages in inappropriate sexual behavior (e.g. sexually active, exhibits self, sexual abuse towards family members or others)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	40. My child has difficulty waiting his/her turn in activities or conversations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	41. My child thinks about suicide, says s/he would be better off if s/he were dead	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	42. My child complains of nightmares, difficulty getting to sleep, oversleeping, or waking up from sleep too early	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	43. My child complains about or challenges rules, expectations or responsibilities	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	44. My child has times of unusual happiness or excessive energy.....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	45. My child handles frustration or boredom appropriately.....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	46. My child has fears of going crazy.....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	47. My child feels appropriate guilt for wrongdoing.....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	48. My child is unusually demanding	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	49. My child is irritable.....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	50. My child vomits or is nauseous more than other children of the same age.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	51. My child becomes angry enough to be threatening to others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	52. My child seems to stir up trouble when bored.....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	53. My child is appropriately hopeful and optimistic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	54. My child experiences twitching muscles or jerking movement..... in face, arms, or body	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	55. My child has deliberately destroyed property	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	56. My child has difficulty concentrating, thinking clearly, or attending to tasks	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	57. My child talks negatively, as though bad things were all his/her fault.....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	58. My child has lost significant amounts of weight without medical reason.....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	59. My child acts impulsively, without thinking of the consequences.....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	60. My child is usually calm.....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	61. My child will not forgive her/himself for past mistakes.....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	62. My child lacks energy.....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	63. My child feels that he/she doesn't have any friends, or that..... no one likes him/her	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	64. My child gets frustrated and gives up, or gets upset easily	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Youth Outcome Questionnaire (Y-OQ® 2.01)

Name _____

Date _____

Never or Almost Never Rarely Sometimes Frequently Almost Always or Always

PURPOSE: The Y-OQ® 2.01 is designed to describe a wide range of troublesome situations, behaviors, and moods that are common to adolescents. You may discover that some of the items do not apply to your current situation. If so, please do not leave these items blank but check the "Never or almost never" category. When you begin to complete the Y-OQ® 2.01 you will see that you can easily make yourself look as healthy or unhealthy as you wish. Please do not do that. If you are as accurate as possible it is more likely that you will be able to receive the help that you are seeking.

DIRECTIONS:

- Read each statement carefully
- Decide how true this statement is during the past 7 days.
- Completely fill the circle that most accurately describes the past week.
- Fill in only one answer for each statement and erase unwanted marks clearly.

Developed by
Gawain Wells, Ph.D.,
Gary M. Burlingame, Ph.D. and
Michael J. Lambert, Ph.D.

© Copyright 1996 American
Professional Credentialing
Services LLC.
All Rights Reserved. License
Required For All Uses

For More Information Contact:
AMERICAN PROFESSIONAL
CREDENTIALING SERVICES
LLC
PO Box 970354
Orem, Utah 84097-0354

E-MAIL:
APCS@OQFAMILY.COM

WEB:
WWW.OQFAMILY.COM
TOLL-FREE: 1-888-MH
SCORE, (1-888-647-2673)
FAX: 1-801-434-9730

1. I want to be alone more than other children of the same age.....	<input type="radio"/>				
2. I have headaches or feel dizzy.....	<input type="radio"/>				
3. I don't participate in activities that used to be fun.....	<input type="radio"/>				
4. I argue or speak rudely to others.....	<input type="radio"/>				
5. I have more fears than other my age.....	<input type="radio"/>				
6. I cut classes or skip school altogether.....	<input type="radio"/>				
7. I cooperate with rules and expectations of adults.....	<input type="radio"/>				
8. I have a hard time finishing assignments, or I do them carelessly.....	<input type="radio"/>				
9. I complain about things that are unfair.....	<input type="radio"/>				
10. I have trouble with constipation or diarrhea.....	<input type="radio"/>				
11. I have physical fights (hitting, kicking, biting, or scratching with my family or others my age.....	<input type="radio"/>				
12. I worry and can't get thoughts out of my mind.....	<input type="radio"/>				
13. I steal or lie.....	<input type="radio"/>				
14. I have a hard time sitting still (or I have too much energy).....	<input type="radio"/>				
15. I feel anxious or nervous.....	<input type="radio"/>				
16. I talk with others in a friendly way.....	<input type="radio"/>				
17. I am tense and easily startled (jumpy).....	<input type="radio"/>				
18. I have trouble with wetting or messing my pants or bed.....	<input type="radio"/>				
19. I physically fight with adults.....	<input type="radio"/>				
20. I see, hear, or believe in things that are not real.....	<input type="radio"/>				
21. I have hurt myself on purpose (for example cut, scratched, or attempted suicide).....	<input type="radio"/>				
22. I use alcohol or drugs.....	<input type="radio"/>				
23. I am disorganized (or I can't seem to get organized).....	<input type="radio"/>				
24. I enjoy my relationships with family and friends.....	<input type="radio"/>				
25. I am sad or unhappy.....	<input type="radio"/>				
26. I have pain or weakness in muscles or joints.....	<input type="radio"/>				
27. I have a hard time trusting friends, family members, or other adults.....	<input type="radio"/>				
28. I think that others are trying to hurt me even when they are not.....	<input type="radio"/>				
29. I have threatened to, or have run away from home.....	<input type="radio"/>				
30. My emotions are strong and change quickly.....	<input type="radio"/>				

Youth Outcome Questionnaire (Y-OQ® 2.01)

Name _____ Date _____

	Never or Almost Never	Rarely	Sometimes	Frequently	Almost Always or Always
PURPOSE: The Y-OQ® 2.01 is designed to describe a wide range of troublesome situations, behaviors, and moods that are common to adolescents. You may discover that some of the items do not apply to your current situation. If so, please do not leave these items blank but check the "Never or almost never" category. When you begin to complete the Y-OQ® 2.01 you will see that you can easily make yourself look as healthy or unhealthy as you wish. Please do not do that. If you are as accurate as possible it is more likely that you will be able to receive the help that you are seeking.					
DIRECTIONS:					
<ul style="list-style-type: none"> ▪ Read each statement carefully ▪ Decide how true this statement is during the past 7 days. ▪ Completely fill the circle that most accurately describes the past week. ▪ Fill in only one answer for each statement and erase unwanted marks clearly. 					
<p>Developed by Cawain Wells, Ph.D., Gary M. Burlingame, Ph.D. and Michael J. Lambert, Ph.D.</p> <p>© Copyright 1996 American Professional Credentialing Services LLC. All Rights Reserved. License Required For All Uses</p> <p>For More Information Contact: AMERICAN PROFESSIONAL CREDENTIALING SERVICES LLC PO Box 970354 Orem, Utah 84097-0354</p> <p>E-MAIL: APCS@OQFAMILY.COM</p> <p>WEB: WWW.OQFAMILY.COM TOLL-FREE: 1-888-MH SCORE, (1-888-647-2673) FAX: 1-801-434-9730</p>					
31. I break rules, laws, or don't meet others' expectations on purpose.....	<input type="radio"/>				
32. I am happy with myself.....	<input type="radio"/>				
33. I pout, cry, or feel sorry for myself more than others my age.....	<input type="radio"/>				
34. I withdraw from my family and friends.....	<input type="radio"/>				
35. My stomach hurts or I feel sick more than others my age.....	<input type="radio"/>				
36. I don't have friends or keep friends very long.....	<input type="radio"/>				
37. My parents or guardians don't approve of my friends.....	<input type="radio"/>				
38. I think I can hear other people's thoughts..... or they can hear mine	<input type="radio"/>				
39. I am involved in sexual behavior that my family would not approve of	<input type="radio"/>				
40. I have a hard time waiting for my turn in activities or conversations.....	<input type="radio"/>				
41. I think about suicide or feel I would be better off dead	<input type="radio"/>				
42. I have nightmares, trouble getting to sleep oversleeping, or waking up from sleep too early	<input type="radio"/>				
43. I complain about or question rules, expectations, or responsibilities	<input type="radio"/>				
44. I have times of unusual happiness or excessive energy.....	<input type="radio"/>				
45. I'm generally okay with frustration or boredom.....	<input type="radio"/>				
46. I am afraid I am going crazy.....	<input type="radio"/>				
47. I feel guilty when I do something wrong.....	<input type="radio"/>				
48. I demand a lot from others or I am pushy.....	<input type="radio"/>				
49. I feel irritated.....	<input type="radio"/>				
50. I throw-up or feel sick to my stomach more than others my age.....	<input type="radio"/>				
51. I get angry enough to threaten others.....	<input type="radio"/>				
52. I get into trouble when bored.....	<input type="radio"/>				
53. I'm hopeful and optimistic.....	<input type="radio"/>				
54. Muscles in my face, arms or body twitch or jerk	<input type="radio"/>				
55. I destroy property on purpose.....	<input type="radio"/>				
56. I have a hard time concentrating, thinking clearly, or sticking to tasks	<input type="radio"/>				
57. I get down on myself and blame myself for things that go wrong.....	<input type="radio"/>				
58. I have lost a lot of weight without being sick.....	<input type="radio"/>				
59. I act without thinking and don't worry about what will happen.....	<input type="radio"/>				
60. I am calm.....	<input type="radio"/>				
61. I don't forgive myself for things I've done wrong.....	<input type="radio"/>				
62. I don't have much energy.....	<input type="radio"/>				
63. I feel like I don't have any friends, or that no one likes me	<input type="radio"/>				
64. I get frustrated or upset easily and give up.....	<input type="radio"/>				

APPENDIX D – STUDY FIVE: COLLECTION OF NZ DATA FOR THE OHIO SCALES

- Brief project summary (high school version)
- Information Sheet (primary school version)
- Ohio Scales P form (primary school version)
- Ohio Scales Y form (for high school)

Invitation to participate in a research project

Full details of this project are contained in the attached Information Sheet

All Year 9 – 13 students at [Name of High School] and their parent/caregivers are invited to take part anonymously.

If you decide to participate:

- It will take 5 – 15 minutes
- Parents and students both take part
- There is a very brief questionnaire for you each to complete
 - No one will know who it is about
- Return the questionnaires to school in separate envelopes
by the end of Friday 13 November

This project is part of bigger research to improve the public services offered to families and young people having emotional or behavioural difficulties. By using a questionnaire (like these ones) to see what changes occur for the families, we get an idea of how effective the service is and what improvements need to be made.

This project uses a questionnaire, developed in the USA, to collect information from a NZ community – parents/whānau and young people at [Name of high school]. The answers you give will help us decide whether this questionnaire is okay to use in NZ as it is, or whether more work needs to be done.

We need a large number and wide range of whānau/families to take part.

Thank you very much; we really appreciate you considering this.



MASSEY UNIVERSITY
COLLEGE OF HUMANITIES
AND SOCIAL SCIENCES
TE KURA PŪKENGĀ TANGATA



RESEARCH PROJECT: *Improving Outcome Evaluation in Child, Adolescent & Family Mental Health Services*

INFORMATION SHEET

Introduction

Kia ora! My name is Amber Barry and this project part of my Doctorate in Clinical Psychology. I'm hoping to help improve outcomes for families/whanau by changing the way psychologists and other professionals track the progress of their clients. To help me do this, I am asking for the assistance of parents in the community.

What this project is about

One way to track progress and outcome in services for children and families is to use standardised questionnaires – "outcome measures". We can look for change in the answers between when a whanau/family first come to the service and later on, or when they finish. We can also compare their answers to the average answers of parents of children that age (this is called "normative data"). This can show us whether what we're doing is helping or whether we should try a different approach.

There are lots of these outcome measures, which are mostly created overseas, with overseas normative data. This makes it difficult to compare the answers of New Zealand parents/caregivers, because we're not sure whether our families/whanau generally answer differently to people in other countries. If New Zealand parents/caregivers do answer differently, we'll know that using the overseas normative data is not appropriate and we should collect our own NZ data. The outcome measures most commonly used in New Zealand don't have NZ normative data, so this project is a first step toward gathering information about our tamariki/children.

I'm looking for around 100 parents/caregivers to complete a short outcome measure for this project. It's called the *Ohio Youth Problems, Functioning and Satisfaction Scales* (the "Ohio Scales") and is used with all children (aged 5–18 years) and families in mental health services across Ohio, USA. It has good research behind it and it could be useful for a range of services here too.

I'd like to invite you to participate in this project, and I am very grateful for your time in doing this.

Who can take part

Your school has kindly agreed to help me by distributing these packs to all students at the school. It's important for you to know that your participation is voluntary and anonymous – neither the school nor I will collect names or other identifying details of anyone who takes part.

All families/whanau can take part in this project, as long as your tamariki/child is between 5 and 12 years old and you can read English.

You don't have to be (or have previously been) clients of a mental health service. I want to know about children/tamariki in the general community, so welcome participants from all backgrounds.

INSTRUCTIONS

We estimate filling out the *Ohio Scales* will take you 5 to 10 minutes.

Parents/Caregivers/Whanau members are asked to...

1. Fill out the demographic questions at the top of the form.
2. Circle the answers that best describe things for your tamariki/child.

Don't forget there are two sides to the form!

3. CHECK to make sure no questions have been *accidentally* missed (you have the right to not answer any particular question).
4. SEAL the form in the envelope provided.
5. **RETURN THE ENVELOPE TO THE COLLECTION BOX AT THE SCHOOL OFFICE BY THIS FRIDAY**

Past research suggests that participants are unlikely to experience harm from taking part in this type of project. However, if this project raises issues for your family/whanau, you might like to use the *Personal Help Services* page in the phone book, or the *Yellow Pages* for private counselors and psychologists, who could help. For moderate to severe concerns we encourage you to speak with your GP or contact the *Child, Adolescent & Family Mental Health Service* at the Palmerston North Hospital.

I will collect the envelopes from your school. The answers from all parents will be combined together so we can compare with the combined answers from parents in Ohio, USA. All information you give will be stored securely in a locked filing cabinet at Massey University. The forms will be destroyed at the end of this research. You are welcome to contact me directly if you would like a summary of the findings.

You are under no obligation to accept this invitation. *Completion and return of the Ohio Scales implies consent for your answers to be used for this project. You have the right to decline to answer any particular question.*

If you have any questions or concerns about this project, please contact me at any stage. Thank you very much for considering taking part.

Kind regards

Amber Barry
Researcher
School of Psychology
Massey University
Palmerston North

06 356 9099, extn 7357
A.J.Barry@massey.ac.nz

Prof. Ian Evans and Dr Shane Harvey
Research Supervisors
School of Psychology
Massey University
Palmerston North

04 801 5799, extn 62125
I.M.Evans@massey.ac.nz
S.T.Harvey@massey.ac.nz

This project has been evaluated by peer review and judged to be low risk. Consequently, it has not been reviewed by one of the University's Human Ethics Committees. The researchers named above are responsible for the ethical conduct of this research.

If you have any concerns about the conduct of this research that you wish to raise with someone other than the researchers, please contact Professor Sylvia Rumball, Assistant to the Vice-Chancellor (Research Ethics), telephone 06 350 5249, email humanethics@massey.ac.nz.

THIS IS A VOLUNTARY PROJECT

RETURN OF THE COMPLETED FORM IMPLIES CONSENT FOR THIS
INFORMATION TO BE USED FOR RESEARCH PURPOSES ONLY

Page One

Ohio Youth Problem, Functioning, and Satisfaction Scales **P**

Parent Rating – Short Form

Child's age: _____

Child's gender: Male Female

Year at school: _____

Ethnicity child identifies with: _____

Form completed by:

 Mother Father Step-mother Step-father Other: _____

Instructions: Please rate the degree to which your child has experienced the following problems in the past 30 days.	Not at All	Once or Twice	Several Times	Often	Most of the Time	All of the Time
1. Arguing with others	0	1	2	3	4	5
2. Getting into fights	0	1	2	3	4	5
3. Yelling, swearing, or screaming at others	0	1	2	3	4	5
4. Fits of anger	0	1	2	3	4	5
5. Refusing to do things teachers or parents ask	0	1	2	3	4	5
6. Causing trouble for no reason	0	1	2	3	4	5
7. Using drugs or alcohol	0	1	2	3	4	5
8. Breaking rules or breaking the law (out past curfew, stealing)	0	1	2	3	4	5
9. Skipping school or classes	0	1	2	3	4	5
10. Lying	0	1	2	3	4	5
11. Can't seem to sit still, having too much energy	0	1	2	3	4	5
12. Hurting self (cutting or scratching self, taking pills)	0	1	2	3	4	5
13. Talking or thinking about death	0	1	2	3	4	5
14. Feeling worthless or useless	0	1	2	3	4	5
15. Feeling lonely and having no friends	0	1	2	3	4	5
16. Feeling anxious or fearful	0	1	2	3	4	5
17. Worrying that something bad is going to happen	0	1	2	3	4	5
18. Feeling sad or depressed	0	1	2	3	4	5
19. Nightmares	0	1	2	3	4	5
20. Eating problems	0	1	2	3	4	5

CONTINUED OVER THE PAGE...

THIS IS A VOLUNTARY RESEARCH PROJECT
RETURN OF THE COMPLETED FORM IMPLIES CONSENT FOR THIS
INFORMATION TO BE USED FOR RESEARCH PURPOSES ONLY

Page Two

If you are unsure about an answer, please rate it as "3, OK". This might apply to questions 3 and 13, in particular.

Instructions: Please rate the degree to which your child's problems affect his or her current ability in everyday activities. Consider your child's current level of functioning.	Extreme Troubles	Quite a Few Troubles	Some Troubles	OK	Doing Very Well
1. Getting along with friends	0	1	2	3	4
2. Getting along with family	0	1	2	3	4
3. Dating or developing relationships with boyfriends or girlfriends	0	1	2	3	4
4. Getting along with adults outside the family (teachers, principal)	0	1	2	3	4
5. Keeping neat and clean, looking good	0	1	2	3	4
6. Caring for health needs and keeping good health habits (taking medicines or brushing teeth)	0	1	2	3	4
7. Controlling emotions and staying out of trouble	0	1	2	3	4
8. Being motivated and finishing projects	0	1	2	3	4
9. Participating in hobbies (baseball cards, coins, stamps, art)	0	1	2	3	4
10. Participating in recreational activities (sports, swimming, bike riding)	0	1	2	3	4
11. Completing household chores (cleaning room, other chores)	0	1	2	3	4
12. Attending school and getting passing grades in school	0	1	2	3	4
13. Learning skills that will be useful for future jobs	0	1	2	3	4
14. Feeling good about self	0	1	2	3	4
15. Thinking clearly and making good decisions	0	1	2	3	4
16. Concentrating, paying attention, and completing tasks	0	1	2	3	4
17. Earning money and learning how to use money wisely	0	1	2	3	4
18. Doing things without supervision or restrictions	0	1	2	3	4
19. Accepting responsibility for actions	0	1	2	3	4
20. Ability to express feelings	0	1	2	3	4

Thank you very much for your time and willingness to assist with this project.

Please CHECK to make sure you have answered all the questions you wish to.

SEAL THIS SHEET in the envelope provided.

**RETURN the envelope to the collection box at the school office
by this Friday.**

THIS IS A VOLUNTARY RESEARCH PROJECT
RETURN OF THE COMPLETED FORM IMPLIES CONSENT FOR THIS
INFORMATION TO BE USED FOR RESEARCH PURPOSES ONLY

Page One

Ohio Youth Problem, Functioning, and Satisfaction Scales Y

Youth Rating – Short Form (Ages 12-18)

Your age: _____ Your gender: Male Female

Year at school: _____

Instructions: Please rate the degree to which you have experienced the following problems in the past 30 days.	Not at All	Once or Twice	Several Times	Often	Most of the Time	All of the Time
1. Arguing with others	0	1	2	3	4	5
2. Getting into fights	0	1	2	3	4	5
3. Yelling, swearing, or screaming at others	0	1	2	3	4	5
4. Fits of anger	0	1	2	3	4	5
5. Refusing to do things teachers or parents ask	0	1	2	3	4	5
6. Causing trouble for no reason	0	1	2	3	4	5
7. Using drugs or alcohol	0	1	2	3	4	5
8. Breaking rules or breaking the law (out past curfew, stealing)	0	1	2	3	4	5
9. Skipping school or classes	0	1	2	3	4	5
10. Lying	0	1	2	3	4	5
11. Can't seem to sit still, having too much energy	0	1	2	3	4	5
12. Hurting self (cutting or scratching self, taking pills)	0	1	2	3	4	5
13. Talking or thinking about death	0	1	2	3	4	5
14. Feeling worthless or useless	0	1	2	3	4	5
15. Feeling lonely and having no friends	0	1	2	3	4	5
16. Feeling anxious or fearful	0	1	2	3	4	5
17. Worrying that something bad is going to happen	0	1	2	3	4	5
18. Feeling sad or depressed	0	1	2	3	4	5
19. Nightmares	0	1	2	3	4	5
20. Eating problems	0	1	2	3	4	5

CONTINUED OVER THE PAGE...

THIS IS A VOLUNTARY RESEARCH PROJECT
RETURN OF THE COMPLETED FORM IMPLIES CONSENT FOR THIS
INFORMATION TO BE USED FOR RESEARCH PURPOSES ONLY

Page Two

Instructions: Below are some ways your problems might get in the way of your ability to do everyday activities. Read each item and circle the number that best describes your current situation.	Extreme Troubles	Quite a Few Troubles	Some Troubles	OK	Doing Very Well
1. Getting along with friends	0	1	2	3	4
2. Getting along with family	0	1	2	3	4
3. Dating or developing relationships with boyfriends or girlfriends	0	1	2	3	4
4. Getting along with adults outside the family (teachers, principal)	0	1	2	3	4
5. Keeping neat and clean, looking good	0	1	2	3	4
6. Caring for health needs and keeping good health habits (taking medicines or brushing teeth)	0	1	2	3	4
7. Controlling emotions and staying out of trouble	0	1	2	3	4
8. Being motivated and finishing projects	0	1	2	3	4
9. Participating in hobbies (baseball cards, coins, stamps, art)	0	1	2	3	4
10. Participating in recreational activities (sports, swimming, bike riding)	0	1	2	3	4
11. Completing household chores (cleaning room, other chores)	0	1	2	3	4
12. Attending school and getting passing grades in school	0	1	2	3	4
13. Learning skills that will be useful for future jobs	0	1	2	3	4
14. Feeling good about self	0	1	2	3	4
15. Thinking clearly and making good decisions	0	1	2	3	4
16. Concentrating, paying attention, and completing tasks	0	1	2	3	4
17. Earning money and learning how to use money wisely	0	1	2	3	4
18. Doing things without supervision or restrictions	0	1	2	3	4
19. Accepting responsibility for actions	0	1	2	3	4
20. Ability to express feelings	0	1	2	3	4

Thank you very much for giving your time to help with this project

Please CHECK to make sure you have answered all the questions you want to.

Then fold this sheet, PUT IT IN ONE OF THE ENVELOPES, SEPARATE from your parent/caregiver's sheet, and seal it closed.

RETURN both sealed envelopes to the collection box at the school office by Friday XX November 2009.

APPENDIX E – STUDY SIX: INFORMANT FEEDBACK ON THE APPROPRIATENESS OF THE OHIO SCALES FOR NZ

- Information Sheet (Parent/Caregiver version)
- Consent Form (Adult version)
- Feedback Questions Sheet (Young person version)



MASSEY UNIVERSITY
COLLEGE OF HUMANITIES
AND SOCIAL SCIENCES
TE KURA PŪKENGĀ TANGATA



RESEARCH PROJECT: Improving Outcome Evaluation in Child, Adolescent & Family Mental Health Services

The Ohio Scales in NZ: Informant (Parent/Caregiver) Feedback INFORMATION SHEET

Researcher Introduction

My name is Amber Barry and this project is one of several that make up the research part of my Doctorate in Clinical Psychology. I'm hoping to help improve outcomes for families/whanau by changing the way psychologists and other professionals monitor the progress of their clients. To help me do this, I am asking for the assistance of parents and teenagers from the Palmerston North area.

Project Description and Invitation

One way to monitor progress and outcome in services for young people is to use questionnaires to get a 'snapshot' of how things are going. Questionnaires designed for this purpose are called outcome measures and there are many available that are mostly developed overseas, often in the USA or UK. The outcome measures most commonly used in New Zealand services are not adapted to suit our local context. In addition, the most widely used outcome measure in New Zealand Child and Adolescent Mental Health services does not allow for the parent/caregiver or young person's perspective on progress. The *Ohio Youth Problems, Functioning and Satisfaction Scales* (the "Ohio Scales") appears to be a promising alternative. It is used with all young people (aged 5–18 years) in mental health services across Ohio, USA, and has good support in the research and clinical literature. My research so far suggests it might be useful for a range of services in New Zealand too, and this project will add to my previous findings about its appropriateness for use with families/whanau here.

I'm looking for between five and eight parents/caregivers with children aged between 5 and 18 years for this project. I am seeking brief feedback about the suitability of the language and content of the *Ohio Scales* for use in New Zealand.

I would like to invite you to participate in this project, and I'm very grateful for your time in doing this.

Participants

You have been invited to take part in this project because you were identified by a friend/family member as a parent/caregiver familiar with New Zealand society, language use, and culture. It is important for you to know that participation is voluntary and confidential.

Project Procedures

If you take part, reading the *Ohio Scales*, considering the questions, and providing feedback should take between 5 and 15 minutes, in total.

Participants are asked to do the following:

1. Contact me to indicate you wish to take part – by email, phone or in person.
2. You will be given a copy of the *Ohio Scales* (one double-sided, 48 item questionnaire) and feedback questions and will need to **read through** these.
3. **Consider** the feedback questions.
4. If you have **feedback**, you can either: make written notes about your ideas for me, or I can discuss it with you, using the questions as a guide. If I take notes of your feedback, they will be summarised and returned to you for any corrections.

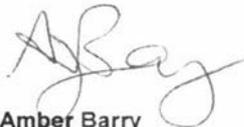
The information from individual participants will not be identifiable when reported as all feedback from parents/caregivers will be collated. All information supplied by participants will be stored securely in a locked filing cabinet at Massey University until they are destroyed at the conclusion of this research. You are welcome to contact me directly if you would like a summary of the findings.

You are under no obligation to accept this invitation. If you decide to participate, you have the right to:

- *decline to answer any particular question;*
- *withdraw from the study at any time;*
- *ask any questions about the study at any time during participation;*
- *provide information on the understanding that your name will not be used unless you give permission to the researcher;*
- *be given access to a summary of the project findings when it is concluded.*

If you have any questions or concerns related to this project, please contact me or my supervisors at any stage. Thank you very much for considering taking part in this project.

Kind regards



Amber Barry
Researcher
School of Psychology
Massey University
Palmerston North

06 356 9099, extn 7357
A.J.Barry@massey.ac.nz

Prof. Ian Evans and Dr Shane Harvey
Research Supervisors
School of Psychology
Massey University
Palmerston North

04 801 5799, extn 62125
I.M.Evans@massey.ac.nz
S.T.Harvey@massey.ac.nz

This project has been evaluated by peer review and judged to be low risk. Consequently, it has not been reviewed by one of the University's Human Ethics Committees. The researchers named above are responsible for the ethical conduct of this research.

If you have any concerns about the conduct of this research that you wish to raise with someone other than the researchers, please contact Professor Sylvia Rumball, Assistant to the Vice-Chancellor (Research Ethics), telephone 06 350 5249, email humanethics@massey.ac.nz.



MASSEY UNIVERSITY
COLLEGE OF HUMANITIES
AND SOCIAL SCIENCES
TE KURA PŪKENGĀ TANGATA



***RESEARCH PROJECT: Improving Outcome Evaluation in Child,
Adolescent & Family Mental Health Services***

**The Ohio Scales in NZ: Informant (Adult) Feedback
PARTICIPANT CONSENT FORM**

I have read the Information Sheet and have had the details of the project explained to me. My questions have been answered to my satisfaction and I understand that I may ask further questions at any time.

I agree to participate in this project under the conditions set out in the Information Sheet.

Signature: _____ Date: _____

Full name – Print: _____



MASSEY UNIVERSITY
COLLEGE OF HUMANITIES
AND SOCIAL SCIENCES
TE KURA PŪKENGĀ TANGATA



**RESEARCH PROJECT: Improving Outcome Evaluation in Child,
Adolescent & Family Mental Health Services**

**The Ohio Scales in NZ: Informant (Youth) Feedback
FEEDBACK QUESTIONS**

Thanks for reading through the *Ohio Scales*. I'm looking for feedback from you about anything you think should be changed for the questions to make sense to New Zealand young people.

I don't need feedback about what things it asks about, or the design of the questionnaire because other people have already talked to me about that. What I want you to do is tell me if there are any ideas or words that would be okay for young people in the USA (where the questionnaire comes from), but might not suit young people here in NZ. I'd be glad to hear any comments you've got – it doesn't matter if you notice lots of things or none at all!

To give feedback in a way that suits you best, you can either,

- **Talk with me** about your feedback (in person/by phone), and I'll write down your ideas,
- **Write answers under the questions** on this sheet, and return it to me,
- **Write/draw on the actual *Ohio Scales* sheet** to show your feedback, and return it to me,
- Or, a combination of these.

Here is what I'm interested in, so have a look at the *Ohio Scales* and think about these things...

1. Do any changes need to be made so New Zealand English is being used (maybe spelling, grammar, or punctuation)? What do you suggest?
2. How should the wording be changed to suit both Maori and Pakeha young people?
3. Do the questions describe things in ways that are different to how we'd describe them in NZ? Can you give examples?
4. Are there ideas or language that is very "USA", but not very "NZ"?
5. Anything else that doesn't fit in NZ like it does in the USA?

Thank you for your help with this project! It is much appreciated – Amber 😊