The Usefulness of the MAYSI-2 and the Mental Health Needs of Youth Offenders in Youth Justice Facilities in New Zealand

Sean Allan McArdle

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Clinical Psychology

The University of Auckland
2018

Abstract

Youth offenders in secure care facilities present with high rates of mental health and behavioural difficulties. An accurate assessment of the needs of this vulnerable group is required for referral to services and access to treatment. A common pathway for assessment is through the use of psychometric measures which look at mental health difficulties and alert staff to youth who may need further assessment and possible treatment.

This research had two aims: to investigate the usefulness the Massachusetts Youth Screening Instrument-Second Version (MAYSI-2), a commonly used screening tool within the New Zealand (NZ) youth justice system; and to provide a better understanding of the mental health needs of youth offenders in secure care facilities.

The sample consisted of 527 youth admitted to one of four secure care facilities within NZ between 2014 and 2016 who completed the MAYSI-2 and a number of other psychometrics.

Confirmatory Factor Analysis revealed issues specifically relating to the validity and reliability of the MAYSI-2 within our sample. Mental health needs of the sample were unique. Youth displayed high levels of substance use difficulties, relatively low levels of psychological distress, and similar rates of suicide risk compared with international studies. Findings are discussed with relation to the international literature on both the MAYSI-2 and wider psychopathology of youth offenders in secure care facilities. Clinical implications with respect to the broader use of psychometrics within NZ's unique youth offender populations are also discussed.

Acknowledgements

First and foremost, to those young people whose inner most experiences are written about in this thesis; I thank you, sincerely.

I would like to thank my primary supervisor Associate Professor Ian Lambie. His insights, suggestions, and support were invaluable contributions to the completion of this thesis. I am deeply appreciative of the time, expertise, and encouragement he has shown to this project and to my professional development as a clinician.

A special thank you to Denise Tapper, Clinical Manager, Oranga Tamariki for all your time and support. Without you this project would never have been an option. I would also like to acknowledge Shanti Patel and the team at the Grey Lynn Oranga Tamariki for their time, assistance, and trust. It was a pleasure working with you all.

I would like to take this opportunity to mention of a number of people who offered various forms of support throughout my clinical training. Thank you to the staff of the clinical programme – my time here has been much enjoyed. To my previous placement supervisors, Malcolm Wright and Marthinus Bekker, the time, expertise, and encouragement you gave was greatly appreciated. A special thank you to my internship supervisor, Hannes Venter, for being so generous with his time and knowledge. You have shaped my clinical practice and equipped me for a long and fulfilling career as a Clinical Psychologist.

To my classmates; thank you for all of the support you have given me over this four year journey. I could not have picked a better group to have gone through it with.

Thank you to my family. Mum, for your unwavering, support, optimism, and encouragement. Dad, for fostering in me a desire to seek out what I don’t know and the ability to be comfortable with feeling more confused after. To Jenny and Brendan for demonstrating, by example, the importance of doing what you love for a living. To my partner, Hilary, for always being there for me.

Table of Contents

Table of Tables and Figures

Chapter 1. Research Overview and Literature Review

Introduction

It is now well documented that youth offenders in secure facilities have substantially higher rates of psychological and mental health difficulties than age equivalent youth in the general population (Abram, Teplin, McClelland, & Dulcan, 2003; Teplin, Abram, McClelland, Dulcan, & Mericle, 2002). This reality becomes problematic when considering the negative impact emotional and behavioural problems can have on the therapeutic alliance and the effectiveness of treatment interventions aimed to reduce recidivism (Bonta & Andrews, 2007). Receiving treatment for mental health problems is likely to play a key role in reducing the likelihood of future recidivism, and while few studies have looked specifically at this issue, preliminary evidence suggests this is indeed the case (Cuellar, Markowitz, & Libby, 2004; Evans Cuellar, McReynolds, & Wasserman, 2006). Additionally, it is in the public's best interests to attend to mental health disorders that may have played a role in illegal behaviour, as well as those which pose an immediate risk or danger to others or themselves. Resources in the youth justice system are notoriously scarce, and the services that do exist are reliant on the accurate identification of youth in need of treatment. Consequently, the mental health screening tools used by the youth justice system must be accurate, reliable, valid, and have clinical utility.

The research conducted in this thesis intends to assess the usefulness of the Massachusetts Youth Screening Instrument, Version 2 (MAYSI-2; Grisso & Quinlan, 2005) in providing a preliminary assessment of the mental health needs of New Zealand (NZ) youth offenders in secure facilities. It also aims to provide a better understanding of the mental health needs and characteristics of adolescents in secure care facilities within a NZ context.

Mental Health Problems of Youth: Internationally and in New Zealand

A significant portion of youth suffer from mental health problems, and there has been a tremendous growth in studies concerned with finding out exactly how many youth, and from which mental health problems they suffer. International studies consistently report that mental health problems effect between 10-20% of children and adolescents globally (Kieling et al., 2011; World Health Organization, 2000; Polanczyk, Salum, Sugaya, Caye, & Rohde, 2015). Suicide has been identified as the fourth leading cause of death among young males and the third for young females (D. Wasserman, Cheng, & Jiang, 2005), and around 50% of mental disorders suffered by adults have their onset before age 15 (Paus, Keshavan, & Giedd, 2008). What these findings illustrate is that adolescence is a critical period with heightened risk of mental, emotional, and behavioural disorders.

Research looking at adolescent mental health prevalence has been widespread, with studies being conducted in a number of developed and developing countries. A recent and comprehensive meta-analytic review which included 41 studies from 27 countries [including NZ (Anderson, Williams, McGee, & Silva, 1987; Fergusson, Horwood, & Lynskey, 1993)], sought to calculate the prevalence of specific mental health problems in children and adolescents (6-18 years of age) worldwide (Polanczyk et al., 2015). Their findings are as follows; any disorder was 13.4%, which is slightly higher than the 7-12% bracket proposed by previous research (Roberts, Attkisson, & Rosenblatt, 1998). It is likely the age range of participants employed in the latter study of 1-18 years of age resulted in a lower prevalence bracket, given mental health problems in early childhood are less common than in adolescents. Any anxiety disorder was 6.5%, any depressive disorder was 2.6%, attention-deficit hyperactivity disorder was 3.4%, and any disruptive disorder (e.g., Conduct Disorder, Oppositional Defiant Disorder) was 5.7%. This study was the first of its kind concerned with

children and adolescents, and once methodological factors were adequately controlled, showed that there were limited cross-cultural and time-based differences for prevalence rates of mental disorders.

Levels of ‘psychological distress’, in epidemiological studies of youth also provide useful data on population based rates of anxiety and depressive symptoms. The majority of general population adolescents tend to fall in a range deemed to be ‘mentally well’ (Green, Gruber, Sampson, Zaslavsky, & Kessler, 2010; Mewton et al., 2016; Peiper, Clayton, Wilson, & Illback, 2015), with the exclusion of one study, Chan and Fung (2014), which found being ‘mentally well’ was the case for only 44.7% of their sample of Hong Kong adolescents. More consistently, between 70-90% of youth are deemed to be free from psychological distress, with a small minority reporting mild, moderate, and severe levels (Green et al., 2010; Mewton et al., 2016; Peiper et al., 2015). A clear gender difference is also observed in most studies with females commonly reporting higher rates of psychological distress compared to their male counterparts (Drapeau et al., 2010; Mewton et al., 2016).

Given the distinct lack of research conducted in NZ on the mental health needs of youth offenders, the following sections will instead discuss the mental health problems common to NZ youth in the general population.

Of particular importance to the current research are the results from Christie et al. (2007, 2011) study looking at substance use difficulties in community sample youth and youth who presented at a community drug and alcohol clinic in NZ. Important because the tool they used to measure substance use difficulties is the same employed in the current study. Results from the community sample indicated that around 20% required further assessment for alcohol

and drug related problems, and 5% were likely in need of specialist substance use services due to serious substance use problems, compared to 70% of the clinical sample (Christie, 2011; Christie et al., 2010). The same study indicated around 78% of the clinical sample likely had problems which were of a clinical severity likely to require intervention. Youth from the community were on average significantly less likely to have scores indicating a need for further assessment (≥ 2) or intervention (≥ 4) compared with those from the clinical sample (0.9 and 8.6 respectively). Differences in scores by ethnicity in the community sample were present, with Māori youth scoring higher on average than NZ European youth and Pacifika youth (Mean SACS Difficulties Score = 1.6, 1.2, and 1.1 respectively).

New Zealand has the advantage of having two major longitudinal studies which have looked at, among other things, the prevalence of mental health problems through childhood and adolescence. These are the Dunedin Multidisciplinary Health and Development Study (DMHDS) and the Christchurch Health and Development Study (CHDS). While both of these studies are south island based, they likely provide a good proxy for prevalence rates nationally. The CHDS has a birth cohort of 1,265 children, and their families, who were born in the Christchurch urban region in 1977. These children represented 98% of all children born in all maternity units in the Christchurch urban region during 1977. The DMHDS has a birth cohort of 1,037 children (91% of eligible births), who were born in Dunedin between April 1972 and March 1973. Retention rates have been high for both studies and follow ups were conducted at similar intervals throughout childhood and adolescents (5, 9/10, 15, 18, and 21; Fergusson et al., 2003). Both the CHDS and DMHDS findings point to a high rate of mental health problems in NZ youth. Starting as early as 11 years old Anderson et al. (1987) from the DMHDS found the overall prevalence of any disorder in their birth cohort of 792 to be around 18% with alarmingly high comorbidity rates (55%).

Follow-up to adolescence (age 15) indicated relatively high persistence of these mental health problems over time, and an increase in overall prevalence to 22% (McGee et al., 1990). Disorders were more common in 15 year old girls (25.9%; largely for internalising disorders such as anxiety and depression) than boys (18.2%). Results from the CHDS were similar. At age 15, 25% of youth met criteria for at least one DSM-III-R diagnosis, and the authors noted a much higher prevalence of disorders in females than in males (33% and 20%, respectively), especially for anxiety and mood disorders (Fergusson et al., 1993). Prevalence of specific disorders was generally consistent between the DMHDS and CHDS at age 15, anxiety disorders were 10.7% and 13.3% respectively, mood disorders were 4.2% and 6.6%, conduct/oppositional disorders were 9.0% and 10.8%, and attention-deficit disorder was 2.1% and 4.8% (Fergusson, Horwood, & Lynskey, 1997).

By age 18, 36.6% of the Dunedin birth cohort were considered to have a mental health disorder, the most prevalent being major depressive episode (16.7%), alcohol dependence (10.4%) and social phobia (11.1%; Feehan, McGee, Raja, & Williams, 1994). High comorbidity was also noted at 18, with 46% of those with a disorder having two or more (Feehan et al., 1994). Prevalence of any mental health disorder at age 18 from the CHDS cohort was slightly higher at 41.8% (Fergusson et al., 2003), though can be explained by the inclusion of a life-functioning criteria in the DMHDS study which pulled the prevalence down from 44.7% to 36.6% (Feehan et al., 1994). For both studies, gender differences were temporally consistent, with girls reporting higher prevalence of disorders (CHDS: 45% vs 39%; DMHDS: 42.3% vs 31.1%), though this difference failed to reach significance in Christchurch cohort (Fergusson et al., 2003). Both studies also had relatively similar prevalence of specific disorders at age 18, anxiety disorders were 19.7% in the DMHDS and 17.1% in the CHDS,

mood disorders were 18.0% and 22.1%, conduct/oppositional disorders were 5.5% and 4.8%, and substance dependence disorder was 12.2% and 8.6% (Fergusson et al., 1997).

Data from the CHDS showed that Māori youth are at a much higher risk of developing significant mental health problems than non-Māori; 49.5% met the criteria for a formal diagnosis compared with 33.1% of non-Māori (Fergusson et al., 1997). Specifically, Māori youth aged 18 were significantly more likely to experience anxiety disorders (24.2% vs 16.1%), mood disorders (29.7% vs 20.1%), conduct disorder (12.1% vs 3.9%), and substance dependence disorders (16.5% vs 7.8%; Fergusson et al., 1997).

A more recent epidemiological study from Auckland, the Youth2000 study, also aimed to gauge the prevalence of mental health problems faced by New Zealand youth. A nationally representative sample of secondary school students (aged 13-18) were assessed in 2001 (n=9,699), 2007 (n=9,107) and 2012 (n=8,500) using a self-administered health and wellbeing questionnaire consisting of the WHO-5 (measure of wellbeing), Reynolds Adolescent Depression Scale (RADS-SF; measure of depressive symptoms), Strengths and Difficulties Questionnaire (SDQ; emotional symptoms, conduct problems, hyperactivity, peer problems, and a prosocial scale), and a number of specific questions regarding periods of low mood, deliberate self-harm, and suicidality (Fleming et al., 2014). Consistent with previous research, they reported higher 12 month prevalence rates of emotional symptoms among females (Fleming et al., 2014). Female students were found to be consistently higher in measures of depressive symptoms (16% vs. 9%), suicide ideation (21% vs. 10%), having a suicide plan (11.5% vs. 6.1), attempt (6% vs. 2%), and deliberate self-harm (29% vs. 18%) compared to their male counterparts (Clark, Fleming, Bullen, Denny, et al., 2013; Fergusson, Boden, & Hayne, 2011).

Between 2001, 2007, and 2012 students reported less sexual abuse and fighting, but exposure to family violence and depressive symptoms did not improve over this time frame (Clark, Fleming, Bullen, Crengle, et al., 2013). The Youth2000 studies have also looked specifically at differences due to ethnicity. Results from 2007 showed that when compared to NZ European students, Māori students were more likely to report suicidal thoughts (17.4% vs. 12.4%), suicide plans (11.1% vs. 7.4%), and suicide attempts (6.9% vs. 3.6%; Fortune, 2010).

The most recent results from the Youth'12 study found Māori students were more likely to self-report that they had made a suicide attempt and deliberately self-harmed (Crengle et al., 2012), and Māori experiencing significant depressive symptoms had increased from 10.6% to 13.9% between 2007 and 2012 (Crengle et al., 2012). Witnessing domestic violence and physical abuse in the home towards another child was also twice as likely for a Māori youth than an NZ European (17.1% vs 9.2% and 10.3% vs 4.4% respectively; Crengle et al., 2012).

Te Rau Hinengaro: The New Zealand Mental Health Survey (n = 12,992) reported that young people aged 16-24 years have the highest 12 month prevalence of mental health disorders of all age groups in New Zealand, at 28.6% (17.7% any anxiety disorder, 12.7% any mood disorder, 9.6% any substance use disorder; Oakley-Browne, Wells, Scott, Kessler, & Üstün, 2008). The survey also found Māori and Pacifika people (Māori more so) had a higher prevalence of disorders, higher prevalence of serious disorders, and that they were less likely than other groups to access treatment when severity was taken into account (Oakley-Browne et al., 2008).

Summary of findings.

There are a number of conclusions that can be made from the literature concerning the mental health prevalence of youth in NZ. Firstly, it can be said that NZ youth have a relatively high prevalence of psychological disorders when compared to prevalence estimates internationally; global estimates of 13.4% are lower than findings from both the CHDS and DMHDS at each age level investigated. Secondly, consistent with international literature, female youth in NZ display an increased level of psychological morbidity when compared to males, and this pattern was consistent across time, from age 11 to 15 to 18. Thirdly, Māori youth present with greater incidence of mental health problems than other ethnicities in NZ. This finding was observed in the CHDS, the Youth2000 survey, and The New Zealand Mental Health Survey, despite different methodologies and assessment techniques.

Māori and the Youth Justice System

The most up to date statistics reveal that the proportion of youth offenders that identify as Māori has increased from 47% to 62% nationwide in the last 10 years, despite only encompassing about a quarter of the NZ population under the age of 17 (Ministy of Justice, 2016). This overrepresentation of Māori youth in the youth justice sector is long established (see Figure 1). While research is being conducted to increase our understanding of Māori offending and how this can be effectively addressed (Caldwell, 2009), the mental health of these young people has been more of a secondary research area. What we do know is that community samples of young Māori are at a substantially increased risk of developing significant mental health problems when compared to their non-Māori counterparts (Fergusson et al., 1997). Young Māori exposure to traumatic events is almost double that of Pakeha (10.3% vs 4.4% respectively; Crengle et al., 2012). Further, over the 10 years from 2003 to 2012,

Māori youth suicide rates have been at least 1.7 times the non-Māori youth suicide rates (Ministry of Health, 2015) and in 2010 rates of suicide among Māori youth was more than 2.5 times higher. Research into the rates of psychopathology for those in Youth Justice secure care facilities links into this concerted effort to fill the gaps in literature and provide an evidence base from which steps can be taken to address these concerning statistics.

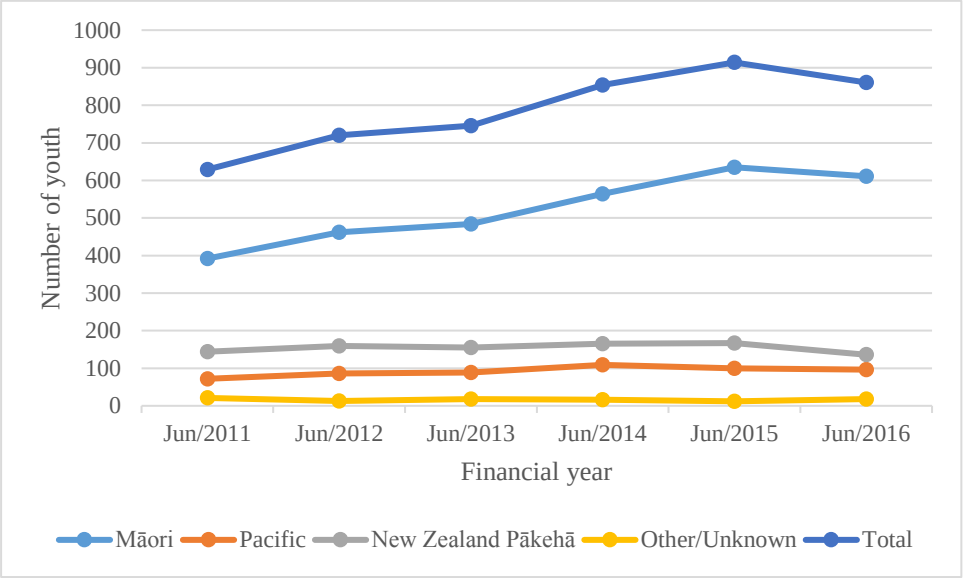


Figure 1. Number of youth admissions to YJ Residences for financial years ending 30 June 2012 to 2016 by ethnicity

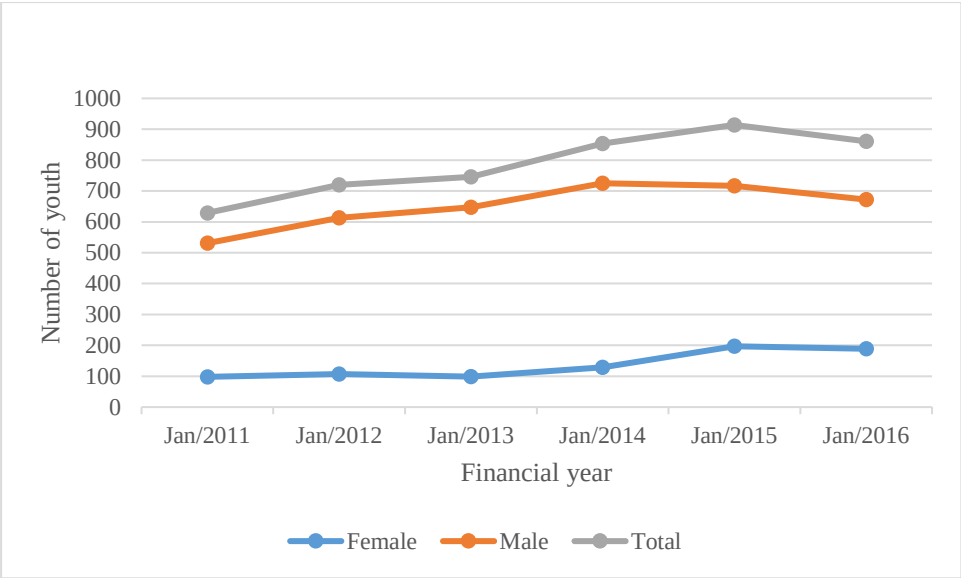


Figure 2. Number of youth admissions to YJ Residences for financial years ending 30 June 2012 to 2016 by gender

Prevalence rates for mental health problems of youth in secure facilities: cause for alarm

Prevalence rates of mental health problems in youth held in secure facilities are significantly higher than comparative community samples. A large number of studies have looked at the combined prevalence of mental health problems for youth offenders. While prevalence has varied from study to study, a large portion found that between 70%-100% of youth in secure facilities met the criteria for at least one mental health diagnosis. This finding has been replicated in Australia (87-100%; Allerton & Champion, 2003; Bickel & Campbell, 2002; Indig, Claudia Vecchiato, Julie Carter, & Natalie Mamone, 2011; Richards, 1996), Austria (90%; Plattner, Aebi, Steinhausen, & Bessler, 2011), Iran (~70%; Ghanizadeh, Nouri, & Nabi, 2012), Canada (92-100%; Gretton & Clift, 2011), the Netherlands (90%; Vreugdenhil, Doreleijers, Vermeiren, Wouters, & Van Den Brink, 2004), Sweden (73%; Ståhlberg, Anckarsäter, & Nilsson, 2010), and England and Wales (95%; Lader, Singleton, & Meltzer, 2003). Studies have also found lower rates of mental health prevalence in detained youth,

though these studies generally look at fewer disorders and hence miss valid diagnosis that may be present (Domalanta, Risser, Roberts, & Risser, 2003; Sawyer et al., 2010). Unsurprisingly, these studies still demonstrated rates substantially higher than community estimates for the disorders they included (Domalanta et al., 2003; Sawyer et al., 2010). Females also tend to have a higher prevalence of mental health problems than males, specifically depression and anxiety (Gretton & Clift, 2011; Teplin et al., 2002). Key studies and their findings referred to throughout this section are displayed in Appendix A, Table 1.

To date, there have been no studies looking at the mental health needs of youth offenders in secure care facilities in NZ. An up to date and comprehensive picture is needed, particularly in light of the steady rates in the Youth2000, CHDS, and DMHDS studies.

Adolescent psychopathology has frequently been dichotomized into two empirically established types: externalizing and internalizing (Forns, Abad, & Kirchner, 2011). Externalizing disorders are characterized by “outward” or external signs of psychopathology, which generate external discomfort and conflict (such as ADHD, Conduct Disorder, substance use, suicide and self harm). Internalizing disorders are characterized by “inward” or internal signs of psychopathology, which generate distress in the individual (such as anxiety, mood disorders, and somatic complaints). Some disorders, such as Post-Traumatic Stress Disorder (PTSD), have been said to have both internalizing and externalizing subtypes (Forbes, Elhai, Miller, & Creamer, 2010; M. Miller & Resick, 2007). The following section will use this categorization system to present the most up to date prevalence estimates for the common externalizing and internalizing disorders that afflict youth offenders in secure care facilities. Differences in prevalence between gender, sex, ethnicity, and country will also be discussed.

Externalizing disorders.

Conduct Disorder/behavioural problems.

Conduct Disorder (CD) is defined as a repetitive and persistent pattern of behaviour in which societal norms and the rights on others are consistently violated, and is characterised by frequent involvement in a range of antisocial activities (American Psychiatric Association, 2013). CD is almost ubiquitous for youth offenders in secure facilities, and for this reason some studies have opted to remove it when calculating the prevalence of psychopathology, given diagnostic criteria are exactly those that tend to warrant secure care (Ståhlberg et al., 2010; Teplin et al., 2002).

Surprisingly, making sense of the studies on CD prevalence, however, is not as straight forward as one may think. A number of studies have, when reporting CD prevalence rates, combined Oppositional Defiant Disorder (ODD) and other behavioural problems into the totals possibly inflating the prevalence rates (G. Wasserman, McReynolds, Schwalbe, Keating, & Jones, 2010). Some studies have also only looked at only males or females, or have combined the two despite there being evidence of gender differences (see Appendix A, Table 1). Prevalence of CD in detained youth has been found to fall anywhere between 35% and 100%, though most international studies are around 60-70% (Archer, Simonds-Bisbee, Spiegel, Handel, & Elkins, 2010; Gretton & Clift, 2011; Hamerlynck, Doreleijers, Vermeiren, Jansen, & Cohen-Kettenis, 2008; Indig et al., 2011; Pliszka, Sherman, Barrow, & Irick, 2000; Richards, 1996; Sawyer et al., 2010; Ståhlberg et al., 2010; Vreugdenhil et al., 2004). Of the larger scale studies conducted (sample sizes greater than 1,000), little firm conclusions can be made. Teplin et al. (2002) found a prevalence rate of 37.8% for males and 40.6% for females with a sample of 1,829 youth. G. Wasserman et al. (2010) found a similarly low rate, with 35.7% of males and females meeting the criteria for a behavioural disorder (CD and ODD) with a sample of

4,961 youth. More recent studies conducted by Archer et al. (2010) and Harzke et al. (2012) found prevalence of CD to be 71.9% and 83.2% respectively; double that found in Teplin et al. (2002) and G. Wasserman et al. (2010). Gender differences of CD prevalence seem to be similarly inconsistent between studies with some finding males have a much higher prevalence than females; 74.6% compared to 45.5% respectively in Archer et al. (2010), and others finding females with a higher prevalence than males; 84.3% and 72.9% respectively in Gretton and Clift (2011) and 40.6% and 37.8% in Teplin et al. (2002). From the perspective of planning and funding mental health services, the implications of prevalence of 36-41% compared to 73-84% are quite profound, and consequently, more research on the issue is needed.

Attention Deficit/Hyperactivity Disorder (ADHD).

ADHD is characterized by the display of developmentally inappropriate levels of inattention and/or hyperactivity-impulsivity which occurs across a variety of settings such as home, school/work, and with peers, and interferes with the person's functioning or development (American Psychiatric Association, 2013). ADHD has been established in the literature as being associated with delinquent behaviour and conduct problems (Retz et al., 2004), and a diagnosis of ADHD makes a young person between four and five times more likely than controls to have been arrested for serious antisocial behaviour (Satterfield, Swanson, Schell, & Lee, 1994). While its presence as a risk factor for offending is well recognised, it's prevalence in detained youth is harder to determine. Studies have ranged from 1% (Atkins et al., 1999) to 76% (Timmons-Mitchell et al., 1997), with most falling between 10% and 30% (Allerton & Champion, 2003; Gretton & Clift, 2011; Hamerlynck et al., 2008; Indig et al., 2011; Robertson, Dill, Husain, & Undesser, 2004), including those with larger sample sizes (Harzke et al., 2012; Teplin et al., 2002). Differences between females and males within studies suggest females had higher rates of ADHD with the exception of one study (Timmons-Mitchell

et al., 1997). This finding may be somewhat surprising as ADHD is generally diagnosed more frequently in males (American Psychiatric Association, 2013). Furthermore, in light of this finding, given most of the samples which did not calculate gender specific prevalence rates are heavily skewed in favour of males (around 10 to 1) it seems possible pooled prevalence rates of ADHD might be higher if the genders were more equally represented.

Suicide and self-harm.

Suicide and self-harm refer to behaviours involving deliberate attempts to injure or inflict death upon oneself. As the leading cause of death in youth justice facilities, suicide has been subject to much study and attention (McCoy, 2015). In an attempt to identify those at risk and in need of intervention, specific attention has been paid to risk factors and characteristics for adolescent suicidal behaviour (L. Hayes, 2009). The literature most commonly reports on previous suicide attempts (see Stokes, McCoy, Abram, Byck, & Teplin, 2015 for a review), but some studies included rates of self-harm (Chitsabesan et al., 2006), ideation (Archer, Stredny, Mason, & Arnau, 2004; Goldstein et al., 2003; Sawyer et al., 2010), and having a plan (Sawyer et al., 2010; refer to Appendix A, Table 1 for percentages). Pooled prevalence for previous suicide attempts were most commonly between 10-20% (Ghanizadeh et al., 2012; Lader et al., 2003; Sawyer et al., 2010; G. Wasserman, McReynolds, Lucas, Fisher, & Santos, 2002), and in the one study with a large sample size (G. Wasserman et al., 2010) 16.3% of their participants had previously attempted. The one sample that reported suicide attempts by females (Dixon, Howie, & Starling, 2004), 46% had a previous attempt and 57% of those had attempted multiple times which is consistent with the literature indicating females attempt more but males are more likely to complete a suicide attempt (Rohde, Seeley, & Mace, 1997). Prevalence rates of suicidal ideation can be separated out by recall period (e.g., current, last week, month, year, and lifetime), which adds

a level of complexity to the literature. Of particular interest for comparisons to the present study is literature concerning current ideation which is not as common as other periods. Only two such studies were found. Prevalence rates of suicidal ideation on admission were 3% (Archer et al., 2004), and 5.5% (Penn, Esposito, Schaeffer, Fritz, & Spirito, 2003). Confusingly, a number of studies report "current" suicidal ideation but assess over a two week or one month timeframe (see Esposito & Clum, 2002; Plattner et al., 2007).

Substance use/misuse disorders.

It is well documented that youth in secure facilities have high levels of substance use problems. Evidence for this finding has been replicated in the United States (US) a number of times (Archer et al., 2010; Harzke et al., 2012; McClelland, Elkington, Teplin, & Abram, 2004), the Netherlands (Hamerlynck et al., 2008), Australia (Allerton & Champion, 2003; Indig et al., 2011), Sweden (Ståhlberg et al., 2010), Canada (Gretton & Clift, 2011), and England and Wales (Lader et al., 2003). Prevalence estimates of the large scale studies indicate around 50-60% of all youth in secure facilities have a substance use disorder (Archer et al., 2010; McClelland et al., 2004; Teplin et al., 2002; G. Wasserman et al., 2010). Gender differences seem to be moderate and the majority of studies have found males to have a higher prevalence than females (Archer et al., 2010; Domalanta et al., 2003; Lader et al., 2003; Teplin et al., 2002; Timmons-Mitchell et al., 1997).

Internalizing disorders.

Affective/mood disturbances.

Mood disorders are mental health disorders broadly characterized by periods of depression or extreme sadness which cause significant distress and impairment in important areas of functioning (American Psychiatric Association, 2013). Elevated mood is also present in bipolar disorder, which no longer falls under the mood disorder umbrella but was previously considered a mood disorder in the DSM-IV, hence its inclusion under affective/mood disturbances for studies that used DSM-IV for diagnosis. Prevalence of affective and mood disorders in detained youth have been found to range from 6% in the Netherlands (Vreugdenhil et al., 2004) to above 70% in the USA (males=72%, females=88% for depressive disorder; Timmons-Mitchell et al., 1997). Despite this large variation, a substantial portion of studies have been fairly consistent, finding prevalence of any mood disorder to be around 20-40% (Domalanta et al., 2003; Hamerlynck et al., 2008; Indig et al., 2011; Pliszka et al., 2000; Teplin et al., 2002) and prevalence of Major Depressive Disorder to be around 14-30% (Bickel & Campbell, 2002; Indig et al., 2011; Pliszka et al., 2000; Robertson et al., 2004; Ståhlberg et al., 2010; Teplin et al., 2002). These findings are mainly from studies with large sample sizes between 1,000 and 11,000 participants (Archer et al., 2010; Domalanta et al., 2003; Harzke et al., 2012; Teplin et al., 2002; G. Wasserman et al., 2010). All studies which reported prevalence for both males and females found females to have, with varying degrees, higher rates of mood disorders (Archer et al., 2010; Domalanta et al., 2003; Gretton & Clift, 2011; Robertson et al., 2004; Teplin et al., 2002; Timmons-Mitchell et al., 1997). Studies with large enough sample sizes also found statistical significance between male and female prevalence (Archer et al., 2010; Robertson et al., 2004; Timmons-Mitchell et al., 1997).

Anxiety disorders.

Anxiety disorders are generally characterized by an excessive amount of worry and apprehension that interferes with a person's ability to function effectively in everyday life (American Psychiatric Association, 2013). Prevalence of anxiety disorders in detained youth has varied quite substantially in relation to gender and the ethnic composition of participants. One study looking at males in the Netherlands (n=204) showed only 9% had an anxiety disorder of any kind (Vreugdenhil et al., 2004) and another in the US showed 52% of males had an anxiety disorder (Timmons-Mitchell et al., 1997). Within studies, females have been found to have a much higher prevalence of anxiety disorders. Archer et al. (2010) found females were 3 times more likely to suffer from an anxiety disorder than males (20.9% compared with 6.9%). Other international studies reflect this finding (Domalanta et al., 2003; Gretton & Clift, 2011; Robertson et al., 2004; Teplin et al., 2002; Timmons-Mitchell et al., 1997), though generally not to the same extent. While the pattern of female bias is consistent, the prevalence ranges have been harder to get consensus on. The large scale studies (n > 1,000) have varied from 6-8% (Archer et al., 2010; Harzke et al., 2012) to around 20% (Teplin et al., 2002; G. Wasserman et al., 2010).

PTSD and trauma.

It is now widely understood that youth in secure facilities represent a severely traumatized population (Abram et al., 2004; Dierkhising et al., 2013). Studies conducted internationally have all reached the same conclusion; that trauma in this population is more common than not, with findings showing over 50% of youth in secure facilities having experienced significant trauma (e.g., violence related trauma and various forms of child abuse; Abram et al., 2004; Allerton & Champion, 2003; Cauffman, Feldman, Watherman, & Steiner, 1998; Dierkhising et al., 2013; Goldstein et al., 2003; Indig et al., 2011; Ruchkin, Schwab-

Stone, Koposov, Vermeiren, & Steiner, 2002). Studies have also consistently found a gender bias, with the rate of exposure to trauma for females being much higher than that of males (Allerton & Champion, 2003; Cauffman et al., 1998; Indig et al., 2011).

These experiences are the antecedents to developing PTSD, a disorder typified by negative alterations in mood and cognitions, avoidance of trauma related stimuli or situations, hypervigilance, and intrusions (American Psychiatric Association, 2013). PTSD has been associated with impulsivity, risky behaviours (James, Strom, & Leskela, 2014), and anger (McHugh, Forbes, Bates, Hopwood, & Creamer, 2012), and in young people it's course can be chronic, potentially lasting years (Morgan, Scourfield, Williams, Jasper, & Lewis, 2003). Prevalence of PTSD for youth in secure facilities has been calculated to fall anywhere between 1.7% for males (Gretton & Clift, 2011) to 48.9% for females (Cauffman et al., 1998), though most studies have found around 10-20% of all youth in secure facilities to suffer from PTSD (Abram et al., 2004; Allerton & Champion, 2003; Chitsabesan et al., 2006; Indig et al., 2011). However, pooled prevalence may not be the most useful tool as gender differences are not highlighted. On closer inspection, rates of PTSD have been found to be consistently higher for females than for males (Archer et al., 2010; Cauffman et al., 1998; Dembo, Williams, & Schmeidler, 1993; Robertson et al., 2004) which could be due to a number of reasons. It is possible that this bias reflects a difference in the nature of abuse experienced by females and males. Females more commonly report being the victim of sexual or physical abuse, and males more commonly report witnessing a violent event (Allerton & Champion, 2003; Cauffman et al., 1998; Indig et al., 2011). The gender difference could also reflect the increased exposure to trauma by females, or alternatively be a product of the different ways in which males and females respond to different traumas.

Comorbid disorders.

Comorbidity of mental health disorders is more common than not for youth in secure facilities, further illustrating the high burden of psychological illness in this population. Findings across studies indicate that co-occurrence of a psychological disorder paired with a substance use disorder is generally the most common form of comorbidity among this group (Abram et al., 2003; Teplin et al., 2002; G. Wasserman et al., 2002). The nature of comorbid disorders is such that even after youth exit secure facilities, prevalence rates of comorbidity remain high (Abram et al., 2015). A number of studies have also found youth with comorbid conduct and depressive symptoms are particularly vulnerable and at an increased risk of peer rejection (Cole & Carpentieri, 1990), school suspensions (Kovacs, Paulauskas, Gatsonis, & Richards, 1988), insomnia and self-injurious behaviour (Simic & Fombonne, 2001), suicidal ideation (Capaldi, 1992), substance use, and unemployment (Capaldi & Stoolmiller, 1999).

Limitations of prevalence studies.

Given the variability of methods, reporting, and sample frames used across studies, direct comparisons and meta-analysis are not feasible nor recommended (Colins et al., 2010). Common methodological limitations inherent in the literature include small sample sizes (Atkins et al., 1999; Bickel & Campbell, 2002; Pliszka et al., 2000; Timmons-Mitchell et al., 1997), looking at any disorder rather than specific disorders (Atkins et al., 1999; Vreugdenhil et al., 2004), and an unhelpfully wide variety of ways in which researchers get diagnosis (e.g., self-report psychometrics, diagnostic screening tools, clinical interviews). Assessment time frames also vary from study to study, including lifetime prevalence rates, past year, past month, and present, to name a few (Colins et al., 2010).

A further limitation specific to cross-culture comparisons is that societal factors must be taken into consideration, such as adjudication and detention policies, differential access to

mental health services for youth, varying arrest patterns, and background prevalences of mental health problems, all of which may contribute to the observed heterogeneity in prevalences (Fazel, Doll, & Långström, 2008). Consequently, it becomes problematic to disentangle variations of study methods and the independent effects of culture and social aspects from prevalence estimates.

The above limitations have plagued epidemiological research in many fields, not just mental health. Bird (1996) was the first to note that epidemiological studies of mental disorders have taken two methodological forms which conceptualise and classify disorders. The first is the medical classification system based on the DSM-IV (APA, 2000), or the WHO ICD-10 (World Health Organization, 1992), which guides how diagnoses are established. The second uses empirically-based, standardised instruments to quantify self-reported emotional and behavioural issues. The main difference being how disorders are operationalised and what measures are used to establish ‘caseness’. Bird (1996) goes on to state that if cross-cultural comparisons are to be made, epidemiological study must apply similar methods to large, representative samples. Methodological inconsistencies inherent in the diagnosis approach such as assessment procedure, sample demographics, and interrater reliability of clinical diagnosis can cause substantial variations in results, while the empirical approach allows direct comparisons of problem scores across studies and countries in a standardised way.

Methodological limitations have forced the literature to focus more on narrative reviews to get pooled prevalence rates (Colins et al., 2010). To date there has only been one attempt to conduct a meta-analysis on the prevalence of mental health problems in detained youth worldwide (Fazel et al., 2008). This study included 25 surveys involving 13,778 boys and 2,972 girls with a mean age of 15.6 years. Results indicated high rates of psychological disturbance and disorder; among male adolescents 52.8% were diagnosed with CD (same for

females), 10.6% with major depression (compared with 29.2% for females), 3.3% with psychosis (compared with 2.7%), and 11.7% with ADHD (compared with 18.5%). The study also confirmed through meta-regression analysis that prevalence rates for different disorders are influenced by a large number of factors. Of crucial importance was the type of instrument used (e.g., DISC or other instrument), as some yielded lower prevalence rates for depression, ADHD, and CD. This finding in itself is of considerable concern, given the sheer number of psychometric instruments utilised in the literature. Prevalence of depression was also lower in studies with psychiatrists acting as interviewers even when an instrument was used, which should have ideally ruled out any interviewer subjectivity.

The other factors related to heterogeneity were gender (girls had higher prevalence of depression), mean subject age (studies with older participants had lower prevalences of ADHD and higher prevalences of CD), sample size (larger samples had higher prevalences of ADHD), study origin/location (studies from the US had lower prevalences of ADHD), and sampling scheme (consecutive entrants and complete sample having higher prevalences of ADHD compared to stratified/nonstratified random). Further, because variations in cultural and social aspects could not be disentangled from variations in study methods, little could be concluded as to the independent effects of culture on prevalence estimates.

While the Fazel et al. (2008) study shed light on the mental health issues faced by youth in secure facilities, and also raised important questions in regards to the heterogeneity of studies previously discussed, it had a number of limitations. First, almost 90% of the youth included in the review were from studies conducted in the US. A dominant sample from the US may skew the results to reflect more of the prevalence rate there, rather than a true global estimate. Second, a number of specific disorders that have been established as common in forensic adolescent populations were not included in the study (i.e., ODD, SUD, and PTSD). If these were able to be included, the overall prevalence is likely to have increased. Third, it was

not clear in some studies whether the functional impairment criteria for receiving a diagnosis was included. Functional impairment in conjunction to symptom criteria is required under the DSM system to make a diagnosis, while the ICD does not (Uestuen & Kennedy, 2009). Though, even in the cases which did include functional impairment criteria, it has been noted youth are poor at reporting this (Bird et al., 2000). It is highly likely studies which relied on symptom criteria only would yield higher prevalence rates than those which took into account functional impairment.

The authors do however note that obtained prevalence rates for disorders must be interpreted with caution, suggesting prevalence ranges may be a better and more accurate approach to account for heterogeneity.

Following on from Fazel et al. (2008), and also learning from the complications inherent in a quantitative, meta-analytic approach, researchers have conducted systematic narrative reviews of prevalence rates for youth offenders in detention (Colins et al., 2010) and youth offenders in general (detained, remand, community; Casswell, French, & Rogers, 2012; Otto, Greenstein, Johnson, & Friedman, 1992; Vermeiren, 2003). While these studies also present the consistent finding that youth offenders are significantly more likely to have psychological disorders than youth in the general population, they too note methodological inconsistencies which make cross-study comparison problematic (Colins et al., 2010).

Summary of findings.

The above findings are part of a consensus that show youth in secure facilities have markedly increased levels of psychological morbidity; figures in excess of age-equivalent, general population rates. The prevalence of psychological disorders in the general paediatric and adolescent population has been estimated to be between 7-12% (Roberts et al., 1998),

though a more recent meta-analysis put the estimate at 13.4% (Polanczyk et al., 2015). In comparison, the prevalence of psychological disorders in detained youth appears to be upwards of 70%. Females also tend to have higher rates of mental health problems, and are consistently higher for mood disorders, anxiety disorders, PTSD, and trauma, when compared to males (Refer to Appendix A, Table 1). Furthermore, comorbidity is the rule rather than the exception for youth offenders in secure facilities, illustrating that not only is there increased prevalence of mental health problems within this group, but also within individuals belonging to this group.

Despite obvious concern, there is a distinct lack of research on the prevalence and types of mental health problems present in youth within the NZ youth justice system. We know from international literature that this population is afflicted with significant rates of mental health disorders, and it is unlikely NZ represents an exception. The inadequate screening of such disorders prevents identification and subsequent treatment. If left untreated it is highly likely these problems will worsen over time and increase a young person's likelihood of future criminal reoffending (M. Foster, Qaseem, & Connor, 2004). Studies have documented the enduring nature of mental health disorders in this population. Even after exiting secure facilities, prevalence rates of mental health problems remain substantial and higher than rates in comparable studies of the general population (Abram et al., 2015). Furthermore, it is not uncommon for a second comorbid disorder to emerge after the first disorder (Anthenelli, 2010). This is concerning as it pushes the young person further away from successfully navigating their way into adulthood, and closer to a life afflicted by ongoing mental health issues and possible crime. As well as the goal of reducing reoffending, the financial cost, and enhancing public safety, there is an ethical (and in some places, legal) mandate to respond to the mental and emotional needs of young people in the justice system.

The Massachusetts Youth Screening Instrument-Second Version (MAYSI-2)

Identifying youth in need of mental health services is the first step in the process of reducing the levels of morbidity within this group (Colins et al., 2010). Given the limited resources available in the youth justice system it is important to accurately and systematically screen youth upon entry into secure facilities to identify those with the highest needs. One such tool, the MAYSI-2, was developed specifically to assess the mental health needs of this group. The MAYSI-2 aims to identify youths aged 12 to 17 years old who are at risk for serious mental, emotional, and behavioural disorders and in need of clinical intervention within the youth justice setting. The MAYSI-2 is a 52-item screening tool which is administered within the first few days after admission to a secure facility and consists of seven scales of mental health or behavioural problems. Respondents are asked to circle “yes” or “no” concerning whether the item has been true for them “within the past few months” on six of the scales, and “ever in your whole life” for one scale (refer to Appendix C for MAYSI-2 example items). Face valid items suggested by adolescent mental health experts and staff working in secure facilities for youth were refined based on psychometric analysis of the seven factor analytically derived subscales: Alcohol/Drug Use, Angry-Irritable, Depressed-Anxious, Somatic Complaints, Suicide Ideation, and Thought Disturbance, and Traumatic Experience.

The Alcohol/Drug Use scale consists of eight items which aim to identify youth who are currently using alcohol or drugs to a significant degree and who are at risk of substance abuse and dependence. Cut-offs for all items are displayed in Table 1 below. The nine-item Angry-Irritable scale assesses explicit feelings of anger and vengefulness, as well as a tendency toward frustration, tension, and irritability. The Depressed-Anxious scale (nine-items) intends to identify symptoms of mixed depression and anxiety, including inner turmoil, depressed mood, and different manifestations of anxiety. The six-item Somatic Complaints scale assesses the presence or absence of various bodily aches and pains, and physical symptoms of anxiety. A

five-item Suicide Ideation scale aims to address thoughts and intentions about self-harm, and depressive symptoms that may present an increased risk for suicide. The five-item Thought Disturbances scale (males only) screens for altered perceptions of reality that are associated with psychotic disorders, and symptoms of derealisation that may be an early indication of a psychotic state. The Traumatic Experience scale is different from the other six scales as it screens for traumatic events over the youths entire lifetime, not over the “past few months”. It is also gender specific, and unlike the other scales, has no specific cut-off scores. Its intention is to identify whether a youth has had greater exposure to traumatic events compared to other youths. Other studies have used a cut-off of three on the Traumatic Experience scale as a measure of exposure to significant trauma (Cauffman, 2004; Stathis et al., 2008), and scores of two or three have been shown in one study to identify youth with symptoms consistent with likely full or partial PTSD (Kerig, Moeddel, & Becker, 2011).

The MAYSI-2 was designed to be administered, scored, and interpreted with ease, often by non-clinical personnel. It also has a short time frame, taking approximately 10-15 minutes to complete and two-three minutes to score.

Table 1

MAYSI-2 scales and corresponding cut-off ranges

MAYSI-2 Scale	Normal[a] Range	Caution[b] score range	Warning[c] score range
Alcohol/Drug Use	0-3	4-5	6-8
Angry/Irritable	0-4	5-7	8-9
Depressed-Anxious	0-2	3-5	6-9
Somatic Complaints	0-2	3-5	6
Suicide Ideation	0-1	2	3-5
Thought Disturbance[d]	0	1	2-5
Traumatic Experience[e]	N/A	N/A	N/A

[a] Based on US norms.

[b] Indicates youth has scored higher than approximately two thirds of youth in secure facilities.

[c] Identifies approximately the top 10 percent of youths on a given MAYSI-2 scale.

[d] For males only.

[e] Specific cut-off scores for the Traumatic Experiences scale have not been published.

Scores falling above the caution cut-off suggests that the reported problem falls in the clinically significant range, while those reaching the warning cut-off indicate scores higher than 90% of other youths in the youth justice system (Grisso & Barnum, 2006). Scores are generally used to determine which young people are in need of further assessment, with those reaching cut-off needing the most urgent attention. It is important to note the MAYSI-2 was developed as a triage tool and was not designed to provide specific diagnosis. Similar to other triage tools, its primary purpose is to sort youth into two distinct groups: one group with almost all youth who have a mental health disorder to be assessed further, and another that has as few youth as possible with disorders that need no further assessment (Colins, Grisso, Mulder, & Vermeiren,

2015). This approach results in a high number of false positives, though the number of false negatives is minimised, and youth with mental health disorders are less likely to be missed. It is for this reason that the majority of youth in secure facilities score above cut-off on at least one MAYSI-2 scale (Grisso & Barnum, 2006). Consequently, youth who reach cut-off on several scales are generally more likely to have a mental disorder.

The MAYSI-2 is the most widely used mental health screening tool in youth justice settings around the world (Cruise, Marsee, Dandreaux, & DePrato, 2007) and has been utilised in studies in the United States (Cauffman, 2004), Canada (Gretton & Clift, 2011), Australia (Stathis et al., 2008), Belgium (Devlieger & Verschuere, 2010), Romania (Rosan, Frick, Gottlieb, & Fasicaru, 2015), Switzerland (Dölitzsch et al., 2017; Leenarts et al., 2016), Holland (Colins et al., 2014), and the United Kingdom (Lennox, O'Malley, Bell, Shaw, & Dolan, 2014; refer to Appendix H, Table 1).

Reliability.

A significant number of studies and reviews have examined the reliability/internal consistency, and structure of the MAYSI-2. Several studies have reported the alpha coefficients (an estimate of reliability for psychometric tests) for the MAYSI-2 for overall sample, by gender, and by race (Archer et al., 2004; Demirbaş-Çakır et al., 2015; Ford, Chapman, Pearson, Borum, & Wolpaw, 2008; Grisso, Barnum, Fletcher, Cauffman, & Peuschold, 2001; M. Hayes, McReynolds, & Wasserman, 2005). Cronbach’s alphas have ranged from .53 to .90, with most scales being above .70. These alphas are similar to those found in other more comprehensive measures of adolescent psychopathology (e.g., CBCL, YSR, MMPI-A, MACI; Grisso & Barnum, 2006). However, Colins et al. (2014) found alpha coefficients ≤ .70 for the majority of MAYSI-2 scales in their Dutch sample and only three of the seven scales in a Turkish sample reached the generally agreed upon .70 threshold (Demirbaş-Çakır et al., 2015). Alpha

coefficients across groups (i.e., race, ethnicity, and gender) were primarily the same or similar for studies conducted in the US (Archer et al., 2004; Grisso et al., 2001; M. Hayes et al., 2005), with the exception of the Traumatic Experience and Depressed-Anxious scale in the M. Hayes et al. (2005) study (boys were lower than girls on the voice administration) and the Traumatic Experience scale in Archer et al. (2004; boys were lower than girls).

Test-Retest Reliability.

Test-retest reliability has been investigated in a number of studies, with the retest delays varying from 5.6 days (Grisso et al., 2001) to 15 days (Archer et al., 2004) to 111 days (Cauffman, 2004). Studies have revealed that test-retest changes have been significant for boys and only on certain scales (Depressed-Anxious, Somatic Complaints, and Thought Disturbance) over a short time period (6-12 days; Grisso & Barnum, 2006). Other findings show reliability is greater over shorter delays (r=.60 to.82) than longer delays (r=.27 to .70). It has been noted however, that the MAYSI-2 was designed to identify transient emotions and thoughts “during the past few months”, and not to measure long standing traits or other enduring conditions (Grisso et al., 2012).

Conceptual Structure.

The current MAYSI-2 scales were determined on the basis of exploratory factor analysis of the tool’s 52 items (Grisso & Barnum, 2006; Grisso et al., 2001). The factor analysis calculated the optimal way for the items to be grouped and guided the development of the MAYSI-2 scales. Items within a given scale therefore measure something in common, and those items that did not load sufficiently onto a factor were discarded or revised. This conceptual structure has been successfully replicated in the Archer et al. (2004) study, more recently in Colins et al. (2014), Demirbaş-Çakır et al. (2015), Russell, Marsee, and Ryals Jr

(2017), and another study replicated all but the Depressed-Anxious scale (Ford et al., 2008). It should be noted however that in Colins et al. (2014), the confirmatory factor analysis model did not fully meet formal criteria for a 'good' fit.

Construct Validity.

A number of studies have investigated the concurrence of the MAYSI-2 scales with other conceptually and semantically similar measures of mental and behavioural health constructs. The primary study, conducted by Grisso et al. (2001) was the first to examine the MAYSI-2 scales in relation to other conceptually parallel psychometric measures (i.e., MACI, CBCL, and YSR), based on the construct they intended to measure (e.g., suicide risk or substance use problems). They found in all but one instance, the MAYSI-2 scores correlated better with their conceptually parallel scale than non-parallel scale. The convergent validity of the MAYSI-2 has since been established a number of times on a range of psychometric tools that have conceptually or theoretically similar scales (Archer et al., 2010; Archer et al., 2004; Butler, Loney, & Kistner, 2007; Chapman & Ford, 2008; Demirbaş-Çakır et al., 2015; Grisso & Barnum, 2006; Lennox et al., 2014; Shulman, Bechtold, Kelly, & Cauffman, 2016). In general, the MAYSI-2 scales have been found to have substantial relations with validated tools measuring similar constructs, though it has been noted that more research is needed in this area (Grisso et al., 2012).

Concerns with the MAYSI-2.

While tests investigating the reliability and validity of the MAYSI-2 have generally been on samples of youth offenders in the US and have provided good support for its use, studies conducted in other parts of the world have had mixed results. Preliminary support for its use has been established in Belgium (Devlieger & Verschuere, 2010) and Turkey (Demirbaş-

Çakır et al., 2015), though Australia (Stathis et al., 2008), the UK, (Lennox et al., 2014) the Netherlands (Colins et al., 2014), and some studies in the US have expressed varying levels of support for its use. These studies will be discussed in the following sections.

A string of studies in the US have called into question the ability of the MAYSI-2 to accurately identify African American youth in need of further assessment (Cauffman & MacIntosh, 2006; McCoy, 2010, 2011). Quantitative and qualitative studies have indicated the MAYSI-2 may not be racially sensitive, and suggests African American youth experience mental health in a way that is different from the youth whom the MAYSI-2 was normed on and designed for. Other studies, however, have failed to confirm this result. A meta-analysis conducted by Vincent, Grisso, Terry, and Banks (2008) conducted in the US with a pooled sample of 70,423, noted relatively small discrepancies between ethnicities. Furthermore, Archer et al. (2004) examined the psychometric properties of the MAYSI-2 on a sample of predominantly African American youth in secure facilities and concluded that the results were generally encouraging and promoted its use with African American youth. These conflicting results indicate the need for further examination, especially given the MAYSI-2 is used extensively with African American youth (Jaggers, Young, McKinney, Bolland, & Church, 2013).

Colins et al. (2014) noted, while providing support for the construct validity of MAYSI–2, its effectiveness in identifying youth in need of further investigation differed between ethnic groups. Results suggested that the MAYSI-2 in the Netherlands did not predict disorders as well as it did in the US, likely due to the ethnic make-up of detained youth. Further, the sample in Colins et al. (2014) excluded females, and calls from the authors were

made for future studies to test whether the MAYSI-2 was appropriate to use with girls outside the US, highlighting the lack of research in this area.

Stathis et al. (2008) similarly expressed reservations with recommending the MAYSI-2 as a screening tool for Australian indigenous youth in secure facilities based on their findings. The authors noted how indigenous youth had trouble understanding the concepts in a number of the MAYSI-2 questions. An example used in the study illustrates how indigenous youth may have inadvertently underreported on some scales; many failed to identify volatile substances (i.e., glue, paint, and petrol) as ‘a drug’, which would likely bias responding on a number of the Alcohol/Drug Use scale questions. Difficulties were also noted in relation to a number of questions in the Thought Disturbance scale, Somatic Complaints scale, and questions regarding school attendance. The authors hypothesise that these misunderstandings are a likely explanation for the fact indigenous youth failed to screen for higher rates of mental health problems than non-indigenous youth, which is what the available evidence suggested.

The only study conducted in England and Wales to date (Lennox et al., 2014) found 90% of youth included in the sample reached the caution cut-off for any scale on the MAYSI-2, much higher than other comparable studies (refer to Appendix H, Table 1). In practical terms this meant almost all young people would need further assessment. Taking this finding into consideration, the MAYSI-2 becomes less useful, and it may be more efficient to go straight to a more comprehensive mental health needs assessment for admitted young people. The authors note that custody in England and Wales is used less liberally, and seen as ‘a last resort’, which may explain the high scores on the MAYSI-2, as secure facilities are reserved for youth with the highest needs.

These findings suggest that while on a whole the MAYSI-2 is a promising screening tool, its validity, reliability, and psychometric properties need to be investigated and established in novel contexts such as NZ.

Testing for validity and reliability.

In psychology, validity refers to the extent to which a particular instrument measures what it claims to measure, and how well it achieves this (Anastasi, 1968; Cicchetti, 1994). New measures are often validated against well established and existing measures which are interested in the behaviour or trait under consideration (S. Foster & Cone, 1995). Alternatively, performance on a test (i.e., MAYSI-2) can be compared to other independent measures of observable facts about the behaviour or trait of interest (Cicchetti, 1994; S. Foster & Cone, 1995). For the purpose of this study, we can refer to the MAYSI-2 as being a new tool in so much as it has not been validated here in NZ. The process of validity testing is discussed in greater detail in Chapter 2.

Reliability refers to the accuracy, dependability, consistency, and repeatability of results from a given test (Kaplan & Saccuzzo, 2012). Reliability can take many forms, though often refers to a measure of internal consistency (i.e., Cronbach's Alpha (Cronbach, 1951)), which defines the extent items in a test hang together (Cicchetti, 1994). Stated another way, internal consistency is a measure based on the correlations between different items on the same test (Tavakol & Dennick, 2011). Items within a specific scale should correlate higher than items in a distinct scale, given they aim to measure the same construct. For example, a questionnaire that is designed specifically to measure anxiety might ask people to specify how often they feel unable to relax, worried, and experience shortness of breath. Given these items are meant to all be measuring the same construct, anxiety, people should respond to them consistently. That is, someone who is highly anxious should report they often find it hard to relax, feel out of breath,

and worried. The inverse also holds; people who are not anxious should rarely find it hard to relax, experience breathlessness, and feel worried. The American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) all agree that reliability of test scores are not absolute given sources of error are numerous. They go on to state that for this reason, reliability checks must be employed and reported every time test scores are employed (AERA, APA, NCME, 1999; Urbina, 2014)

From a psychometric perspective, validity and reliability are intrinsically linked. A psychological instrument cannot be valid unless it is reliable, though the reliability of an instrument does not depend on its validity (Tavakol & Dennick, 2011). In other words, score reliability is necessary but not sufficient for determining validity (Urbina, 2014). However, for a test to be deemed appropriate for use, both the validity and reliability must be investigated and established.

Need for current research.

It is critical to examine whether the MAYSI-2 is appropriate for use in NZ for several reasons. First, its use has been questioned in countries outside of the US, some of which could be said to be culturally similar to NZ (i.e., Australia and the UK). Based on this concern, it is unsafe to assume it would be appropriate for NZ youth, and thus should be subject to investigation. Second, normative data for the MAYSI-2 was obtained from predominantly White, African American, and Hispanic youth (Grisso & Barnum, 2006). Secure facilities in NZ deal with youth from diverse ethnic backgrounds which differ significantly from those seen in the US (e.g., Māori and Pacifika people). McCoy (2010) was the first to note culture could affect language interpretation of the MAYSI-2 items. It is unknown whether NZ youth will understand all the questions, or latent meaning of some words, leading to inconsistent

reporting. Third, other mental health psychometrics have been shown to be affected by culture and thus their use in NZ cautioned against (Barker-Collo, 2003), and calls for the validation of currently used measures in NZ have been made (Black, Pulford, Christie, & Wheeler, 2010). Fourth, there are potential differences between the US and NZ's sentencing policy. Similar to the UK, NZ law places emphasis on finding alternative means and actions for dealing with youth offenders, and ultimately aims to keep them out of the justice system (Child, Young Persons and their Families Act, 1989). This approach is likely to impact on the proportion of youth in NZ who may be at or above the MAYSI-2 cut-off scores as secure facilities are reserved for the most high risk youth.

In light of the above, the usefulness and appropriateness of a screening tool such as the MAYSI-2 is expected to vary. The key question is by how much, and whether the variation deems the use of the MAYSI-2 in NZ as appropriate. In order to answer this question, the validity and reliability of the MAYSI-2 must be investigated.

The need to validate the MAYSI-2 in NZ is evident, especially given its current use throughout secure care facilities here, and the questions raised by similar studies conducted internationally. This research will also add to the international literature on the MAYSI-2's applicability to contexts outside the US which is an area in need of further study.

While the current protocol for screening youth entering a youth justice secure facility in NZ is of high standard, improvements can still be made. Determining the usefulness of the MAYSI-2 in NZ has the potential to streamline the triage process and assist government organisations in developing accurate assessment procedures to identify youth at high risk of mental, emotional, and behavioural problems. The nature of the MAYSI-2 allows for good cross-cultural comparisons of scores, as its use is limited to youth within the ages of 12-17 (i.e., participants are similar), and scores are obtained under similar conditions (i.e., upon entry into

a secure facility). By using the same instrument and comparable study methods, differences across international studies will thus largely represent differences in geographical, social, and/or cultural aspects (Polanczyk et al., 2015).

The need for an up to date picture of the mental health concerns of this population is well overdue. Furthermore, meaningful and accurate data on the prevalence of mental health issues experienced by NZ youth in secure facilities will allow the allocation of the juvenile justice system's scarce resources to become more targeted, cost effective, and efficient.

Aims of Research

The primary aim of this study is to assess the appropriateness of the MAYSI-2 in NZ by determining whether the robust psychometric properties of the MAYSI-2 generalise from the US sample to our NZ sample. A secondary aim is to investigate the level of mental, emotional, and behavioural problems present within our sample of youth offenders in secure facilities.

Chapter 2. Method

Research Participants

Data for this study was retrospectively gathered from a non-selected group of youths aged 13 to 17 who were admitted to one of NZ's four youth justice facilities between January 2014 and December 2016.

Youth justice secure care facilities in NZ are reserved for the highest risk recidivist youth offenders. There are four facilities which house 146 beds, catering to both male and female youth. The residences were as follows: Korowai Manaaki (Auckland, n=93), Te Au rere

a te Tonga (Palmerston North, n=225), Te Puna Wai o Tuhinapo (Christchurch, n=8), Te Maioha o Parekarangi (Rotorua, n=203). The low numbers from Te Puna Wai o Tuhinapo were due to systems not being in place to ensure that every youth on intake completed the MAYSI-2.

Archival records with unknown gender, birthdate, admission date, and/or ethnicity were removed from the data set (n=7). Missing data on the MAYSI-2, SACS, K6, and Suicide Risk Practice Tool also resulted in six subjects being left out of the sample. One-hundred-and-forty-six youth were removed from the sample due to the MAYSI-2 being conducted five days after entry to the secure care facility. For adolescents who completed the MAYSI-2 on more than one occasion, reflecting multiple admissions to secure units, only the data from the first admission was included for the purposes of the CFA, internal consistency/reliability, and the analysis for convergent validity given they assume independent data (i.e., a single youth represented one data point/admission). Repeat admissions were recorded separately in order to test the MAYSI-2's test-retest reliability. In total there were 274 re-admissions. In these cases, the first admission with the most complete data profile was selected for inclusion (i.e., no missing data points or demographic information and completion of all psychometrics). There were a number of cases whereby individuals were excluded for more than one reason. This exclusion process resulted in a final sample of 222 (111 pairs) with a six month delay and 46 (23 pairs) with a three month delay.

For inclusion in the convergent validity analysis the youths had to have completed the MAYSI-2, SACS, K6, and Suicide Risk Practice Tool within 5 days of each other (mean=2.27, SD=1.38) and within 5 days of arriving to a youth secure care facility (mean=2.28, SD=1.36). Information about ethnicity, age, admission date, and gender must have also been available.

The same rules applied for inclusion in the CFA, internal consistency analysis, convergent validity analysis, and test-retest analysis.

The final samples are as follows CFA (n=527), internal consistency analysis (n=527), liner regression models and t-test used for the convergent validity analysis (n=524), and test-retest analysis (n=222). In total, data from 529 individuals were used in separate analyses.

Males were significantly over represented (n=434, 82%) compared to females (n=95, 18%). In regards to ethnicity, youth were categorised by their chosen primary ethnicity included on their digital case file located on the Care and Protection; Youth Justice; Residences; Adoption; System (CYRAS). In cases where no primary ethnicity was evident, and the youth identified multiple ethnicities, the standard NZ system for prioritising ethnicity was used (Oakley-Browne et al., 2008): Māori ethnicity was prioritised over Pacifika ethnicity, and Pacifika ethnicity was prioritised over other ethnicities. That is, people who reported their ethnicity as Māori and other ethnic groups were classified as belonging to the Māori ethnic group. People who identified as Pacifika, but were not Māori, were classified as Pacifika regardless of the other ethnicities they may have also reported as identifying with. The vast majority of youth identified as Māori (n=404, 76.4%); (n=55, 10.4%) as Pacifika; (n=67, 12.7%) as European; and (n=3, 0.6%) as Other. A breakdown of demographic information by location is provided in Table 2 below.

Table 2

Demographic make-up of participants by location

Demographic	Category		Total N = 529		Auckland N = 93		Christchurch N = 8		Palmerston North N = 225		Rotorua N = 203	
			N	%	n	%	n	%	n	%	n	%
Gender	Male		434	82	67	72.0	6	75.0	167	74.2	198	97.5
	Female		95	18	26	28.0	2	25.0	58	25.8	5	2.5
Age at intake	13		1	0.2	0	0.0	0	0.0	0	0.0	1	0.5
	14		112	21.2	17	18.3	3	37.5	54	24.0	38	18.7
	15		208	39.3	39	41.9	3	37.5	85	37.8	81	39.9
	16		195	36.9	32	34.4	2	25.0	83	36.9	78	38.4
	17		13	2.5	5	5.4	0	0.0	3	1.3	5	2.5
Ethnicity	NZ Māori		404	76.4	67	72.0	7	87.5	168	74.7	162	79.8
	NZ European		67	12.7	7	7.5	1	12.5	33	14.7	26	12.8
	Pacifika		55	10.4	19	20.4	0	0	21	9.3	15	7.4
		Tongan	*(12)*	*(2.3)*	*(4)*	*(4.3)*	*(0)*	*(0.0)*	*(5)*	*(2.2)*	*(3)*	*(1.5)*
		Samoan	*(13)*	*(2.5)*	*(5)*	*(5.3)*	*(0)*	*(0.0)*	*(6)*	*(2.7)*	*(2)*	*(1.0)*
		Cook Island	*(19)*	*(3.6)*	*(5)*	*(5.3)*	*(0)*	*(0.0)*	*(8)*	*(3.6)*	*(6)*	*(3.0)*
		Pacifika other	*(11)*	*(2.1)*	*(5)*	*(5.3)*	*(0)*	*(0.0)*	*(2)*	*(0.9)*	*(4)*	*(2.0)*
	Other ethnicity		3	0.6	0	0.0	0	0	3	1.3	0	0.0

Note: Due to round-off error, percentages may not add up to 100.0%.

Ethics

Prior to the research taking place, an application for ethical approval was placed with the University of Auckland Human Participants Ethics Committee (UAHPEC). Ethical approval was granted on 22 October, 2015 for a period of three years (Ref. 15830). Ethics approval was also sought and granted on 10 March, 2016 from the New Zealand Ministry of Social Development Ethics Committee for the duration of the project.

Measures

The Massachusetts Youth Screening Instrument - Second Version (MAYSI-2).

The MAYSI-2, developed by Grisso and Barnum (2006) aims to identify youths aged 12 to 17 years old who are at risk for serious mental, emotional, and behavioural disorders and in need of clinical intervention within the juvenile justice setting. The MAYSI-2 is a 52-item screening tool which is administered within the first few days after admission to a secure facility and consists of seven scales of mental health or behavioural problems. Respondents are asked to circle “yes” or “no” concerning whether the item has been true for them “within the past few months” on six of the scales, and “ever in your whole life” for one scale. Caution and warning cut-offs are based on the number of items endorsed in a scale. The primary scales are as follows; Alcohol/Drug Use, Angry-Irritable, Depressed-Anxious, Somatic Complaints, Suicide Ideation, and Though Disturbance (for boys only). The seventh scale, Traumatic Experiences, provides information about the presence of trauma throughout their life. The MAYSI-2 has been shown to have robust psychometric properties in the populations it has been tested on (refer to Chapter 1. for the psychometric properties of the MAYSI-2).

The Substances and Choices Scale (SACS).

The SACS was developed in NZ as a brief alcohol and other drug (AoD) screening and outcome instrument for adolescents (Christie et al., 2007). The SACS has proven utility in the screening and identification of AoD difficulties in NZ youth, specifically Māori and Pacifika, and is suitable for 13-18 year olds. The SACS consists of three sections, with Section B being scored differently to Section A and C. Section A contains 12 items and assesses the frequency and range of illicit drug use. Available responses are “never”, “one a week or less”, “more than once a week”, and “most days or more”. Section C records frequency of tobacco use over the last month, and is scored the same as Section A. The 10 items in Section B assess problems associated with substance use (e.g., attendance at school, driven under the influence, unsafe sex/unwanted sexual experience, arguments, got into serious trouble, consuming alone), and are scored by ascribing one of the following labels to the question: “not true”, “somewhat true”, or “definitely true”. These scores are then given a numerical value ranging from 0 for “not true” to 2 for “definitely true”. Section B thus gives a difficulties score out of 20 which may signal the need for further assessment. A score of 2 or above indicates a need for further assessment; a score of 4 or above indicates problems which are of clinical significance and likely require intervention; a score of 6 or above are indicative of serious problems and may require intervention from a specialist substance use service. The SACS has excellent internal consistency (Cronbach’s alpha) of 0.91 and excellent and good test–retest reliability over a 1 and 3 week period of 0.91 and 0.88 respectively. Consultations with young people during the development phase also led to strong face validity and acceptability of its use among the target demographic. A copy of the SACS is available in Appendix D.

The Kessler Psychological Distress Scale (K6).

The Kessler Psychological Distress Scale (K6) was developed in 1994 for screening populations for non-specific psychological distress (Kessler et al., 2002). The scale uses a summative scoring approach over the 6 questions, all of which focus on anxiety and depressive symptoms experienced in the last 4 weeks. Each question has a five value scale: all of the time, most of the time, some of the time, a little of the time, and none of the time (each is assigned a number from 0 through 4). The maximum score is therefore 24, indicating severe distress, and the minimum score is 0, indicating no distress. People who score between 0 and 7 are likely to be well, or absent of mental illness in the previous 30 days. Those who score between 8 and 12 are likely to have a mild to moderate mental health disorder. Those who score in the 13-24 range are likely to have a severe mental health disorder and are high risk (Kessler et al., 2003; Wang et al., 2007). The K6 has been shown to have excellent internal consistency (Cronbach's alpha=0.89; Kessler et al., 2003), and significant correlations with other similar instruments including the World Health Organisation Disability Assessment Schedule (WHO-DAS; Kessler et al., 2003), and the Comprehensive International Diagnostic Interview – Short Form (CIDI-SF; Kessler et al., 2003). The Kessler scales have been used in large scale epidemiological studies here in NZ (NZMHS; Mental Health Commission, 2011, 2012), and have been established as having good measurement precision in the NZ population (Krynen, Osborne, Duck, Houkamau, & Sibley, 2013). It has also been utilised in both youth (Chan & Fung, 2014; Eno Louden, Skeem, & Blevins, 2013; Mewton et al., 2016; Peiper et al., 2015) and forensic populations (Kubiak, Beeble, & Bybee, 2012). Results from epidemiological studies in both adult and adolescent samples commonly demonstrate females having higher rates of psychological distress than males (Drapeau et al., 2010; Mewton et al., 2016). A copy of the K6 is available in Appendix E.

Suicide Risk Practice Tool.

The Suicide Risk Practice Tool was administered in conjunction with the SACS and K6 and consists of the following questions aimed to identify whether there is any current risk of suicide and if further assessment is required; “how do you see the future?” “Do you ever feel that life is not worth living?” “Have you ever thought you would like to end it all?” If one of these answers is in the positive a full Suicide Risk Assessment is carried out to determine the degree of risk. A copy of the Suicide Risk Practice Tool is available in Appendix F.

The SACS, K6, and Suicide Risk Practice Tool make up a single form used in youth justice facilities, abbreviated to the SKS. The SKS will be referred to throughout the remainder of the study when referring to all three screening tools as one. The SACS, K6, and Suicide Risk Practice Tool will also be referred to in their individual form when appropriate.

Procedure

Data of interest was collected through file auditing of both pen and paper files and the online case management system (CYRAS) at a single Child, Youth, and Family office. On entry to the youth justice secure care facility each youth completed a routine structured intake assessment. The MAYSI-2 formed part of this assessment and was administered to all youth within the first few days after entry (mean=2.28, SD=1.36). The total screening process was conducted over a period of seven days and included a variety of assessment instruments including the SACS, the K6, and Suicide Risk Practice Tool. All youth admitted got a health screen (HEADSS: home, education [i.e., school], activities/employment, drugs, suicidality, and sex), and an appointment with a medical doctor regarding general health. Youth were aware that mental health screening was routine and were not under any legal obligation to participate. The corresponding scoring data was recorded at the time of administration in pen/pencil format,

and then loaded onto CYRAS by youth justice facility staff. Individuals were anonymised (given numerical values) and the data from CYRAS was then input onto a Microsoft Excel spreadsheet. Dichotomous and categorical variables were dummy coded so that a negative response was represented by 0 and a positive response was represented by 1. For categorical variables with more than two categories, for example, variables in which the possible responses included *never, once a week or less, more than once a week,* or *most days or more,* the variable was recoded as 0, 1, 2, or 3, respectively. Analysis was conducted using various statistical software, and these will be made explicit in the following sections. It is also important to note that all analysis involving the Thought Disorder scale was calculated using male participants only, as per MAYSI-2 requirements.

Statistical Analysis

Construct validity.

Construct validity is generally defined as the extent to which a test is actually measuring or assessing what it is reporting to (Cronbach & Meehl, 1955; Strauss & Smith, 2009). Construct validity has been described as subsuming all other types of validity evidence (Strauss & Smith, 2009) and the approaches to test construct validity are numerous. It is therefore advised that construct validity be developed from a number of perspectives and by an accumulation of evidence (Strauss & Smith, 2009). Evidence for the construct validity of a particular test or assessment instrument can be provided by the predictive, concurrent, discriminant and convergent validity, criterion-related validity, content validity, and factor structure (Haynes, Richard, & Kubany, 1995). The following section includes a brief description of the approaches used in the current study to assess the construct validity of the MAYSI-2 in an NZ sample, and then the analysis conducted.

Factor structure.

Confirmatory factor analysis (CFA) is a common statistical method of construct validation via factor structure (Thompson, 2004). A factor is a group of items in a test that measure the same latent variable. The higher an item "loads" onto a factor, the stronger the relationship between the item and the underlying characteristic or trait that the factor represents. The aim of CFA is to determine whether the proposed a priori model (generally developed through exploratory factor analysis [EFA]) is a good fit for the sample at hand.

A conscious decision was made to weigh up the pros and cons of using confirmatory factor analytic techniques over EFA with principal component analysis (PCA), as both have been utilized in the MAYSI-2 literature, the latter far more often. While both methods are similar in that they are methods of structural equation modelling (SEM), they are distinct techniques based on different sets of assumptions and used for different purposes. CFA was chosen mainly for the reason that, among other things, it tests the extent to which a proposed factor structure can be replicated in another sample (Van Prooijen & Van Der Kloot, 2001). Revisiting the aims of the current study, which include testing whether the psychometric properties of the MAYSI-2 hold for our NZ sample, it seems appropriate to use CFA to determine whether the proposed factor structure of the MAYSI-2 matches the results of the current study. Thus, the factor structure of the MAYSI-2 in this sample was assessed using goodness-of-fit results obtained through CFA to determine construct validity and test the hypothesised factor structure. This was done first with the entire sample, and then by gender, to ensure that items were loading correctly onto their respective factors at a subgroup level. CFA was carried out using IBM SPSS Analysis of Moment Structures (AMOS).

Convergent validity.

One of two strategies for demonstrating convergent validity (the other is discussed in the next section under CFA) is determining whether scores on a particular test or measure correlate strongly with the score tests or measures that are theoretically related or that claim to measure a similar construct (Goodwin, 2009). This procedure determines whether measures of constructs that should, in theory, be related to each other are, in fact, observed to be so.

The convergent validity of the MAYSI-2 was investigated by comparing scores on a select number of scales to scores on theoretically and semantically similar psychometric instruments that measure similar underlying constructs. MAYSI-2 subscales compared included Suicidal Ideation with the Suicide Risk Practice Tool, Alcohol-Drug Use with the SACS difficulty Scale, and Depressed-Anxious with the K6. Somatic Complaints, Anger/Irritability, Traumatic Experiences, and Thought Disturbance scales were not tested because they do not pose immediate risk (i.e., compared to suicidal ideation) and youth meeting the caution or warning cut-offs are less prevalent than other scales (i.e., compared to substance use and depressed-anxious). Further, data for semantically and conceptually similar scales to the ones mentioned above were not present to analyse. Convergent validity was assessed using R (R Core Team, 2014).

Discriminant and convergent validity in relation to CFA.

Discriminant validity tests whether concepts or measurements that should not be related are, in fact, unrelated (L. Miller, McIntire, & Lovler, 2011). In other words, discriminant validity is shown when each item correlates weakly with all other constructs except for the one which it is theoretically associated with in the specified model.

While a number of ways exist to assess discriminant validity at the construct level, the Fornell and Larcker (1981) technique was recommended, as it has been described as being a

more stringent test (Farrell & Rudd, 2009; Voorhees, Brady, Calantone, & Ramirez, 2016). The emerging technique, heterotrait-monotrait ratio of correlations (HTMT) was not used given it has not yet been established as a superior method for calculating discriminant validity in covariance-based SEM techniques such as those used in AMOS (Henseler, Ringle, & Sarstedt, 2015). For discriminant validity at the construct level to be supported the square root of the Average Variance Extracted (AVE) value for a latent construct (MAYSI-2 scale), should be greater than the correlation estimate for the two constructs of interest. In other words, the latent construct should explain its items measures better than an alternative construct (Hair, Anderson, Babin, & Black, 2010). AVE varies from 0 to 1, and it represents the ratio of total variance that is due to the latent variable and the variance due to measurement error.

The AVE is also used to determine whether convergent validity is present. Convergent validity is defined as the extent to which a specified set of items for a construct converge or share a high proportion of variance in common (Hair et al., 2010).

The AVE estimate for each latent construct should be at least .50 for convergent validity to be supported at the construct level (Fornell & Larcker, 1981; Hair et al., 2010). If the AVE for a scale is less than .50 this means that over 50% of the variance is due to measurement error, which is larger than the variance captured by the construct itself. AVE was calculated using the following formula (Fornell & Larcker, 1981).

$$AVE = \frac{\sum \lambda_i^2}{\sum \lambda_i^2 + \sum_i \mathrm{var}(\varepsilon_i)} \tag{1}$$

In this formula λ_i is the standardized loading for each observed variable and var(ε_i) is the error variance/measurement error associated with each observed variable. All AVE were calculated by hand and then checked using a calculator based on the formula provided by Fornell and Larcker (1981).

Item level convergent validity in relation to CFA is shown when each measurement item loads strongly onto its theoretical construct/latent variable (Hair et al., 2010). It is specified that all items that load onto their specified factor must be statistically significant, be at least .5 or higher, and ideally .7 or higher (Hair et al., 2010). Thus, convergent validity is evidenced through examining the size of the standardized item factor loadings (item level) and the size of the average variance extracted for each construct (factor level).

Item level assessment of discriminant validity was also carried out as it represents another popular approach for establishing discriminant validity (Henseler et al., 2015), and some texts specify that both procedures (i.e., Fornell and Larker, and item level examination) should be utilized (Gefen & Straub, 2005). This method stipulates that discriminant validity is shown when each item correlates strongly with their theoretically assigned factor and weakly with all other factors (Gefen & Straub, 2005). It has been noted that established thresholds for establishing discriminant validity from structure coefficients do not currently exist (Gefen & Straub, 2005), and therefore the interpretations will not be definitive, and will only support the more stringent Fornell and Larcker (1981) method of discriminant validity. It is important to note that high inter-construct correlations can, in certain situations, cause pronounced spurious correlations between theoretically unrelated items and factors (Bollen, 1989), and thus interpretation of high item-factor correlations will be limited.

To recap, both item level and factor level methods were used to assess convergent and discriminant validity using CFA in order to get the most robust assessment of the MAYSI-2 psychometric properties.

Reliability.

Reliability is the degree to which a measure is free of error and therefore yields accurate and consistent results (Kaplan & Saccuzzo, 2012). Cronbach's alpha (Cronbach, 1951) is one

of the most commonly used measure to estimate the reliability (internal consistency) of a psychometric test. Cronbach's alpha will generally increase as the intercorrelations between items measuring the intended latent construct increase (refer to Chapter 1: testing for reliability and validity for a more detailed explanation of internal consistency).

Cronbach's alpha was used to determine the reliability of each of the seven MAYSI-2 scales for the total sample and for gender. Alpha coefficients for boys and girls was also calculated for Māori, Pacifika, and Pakeha/European ethnicities in order to look at ethnic differences in the measures reliability. In the analysis of reliability, Cronbach's alpha coefficients of .60 to .70 were deemed the lower limit of acceptability (Hair et al., 2010) though it should be noted that many researchers use .70 as a rule of thumb (L. Miller et al., 2011). Internal consistency/reliability as measured by Cronbach's alpha was assessed using IBM SPSS Version 23.

Composite reliability (CR; often called construct reliability) is another approach to estimating the reliability/internal consistency. While Cronbach's alpha is the most commonly used estimate, it has been subject to substantial critique due to its underestimation of reliability in certain situations (Peterson & Kim, 2013; Starkweather, 2012). That being said, it has also been noted different reliability estimates often do not produce dramatically different results (Hair et al., 2010). In light of this, psychometricians have suggested more accurate alternatives such as Jöreskog's rho when using SEM (Revelle & Zinbarg, 2009). However, given all tests of reliability produce estimates, it seems prudent to provide multiple estimates wherever possible. CR is expressed by formula (2) (Raykov, 1997):

$$CR = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i)^2 + (\sum \delta_i)} \tag{2}$$

In this formula λ_i is the standardized loading for each observed variable and δ_i is the error variance/measurement error associated with each observed variable. The rule of thumb applied to CR is that it should be equal to or greater than .7 (Hair et al., 2010), though some have promoted anything greater than .6 as being acceptable (Tseng, Dörnyei, & Schmitt, 2006). Jöreskog's rho (also known as McDonald's Omega) is the measure of CR used (Jöreskog, 1971). All CR were calculated by hand and then checked using a calculator based on the formula provided by Raykov (1997).

Test-retest reliability.

The correlation between two administrations of the MAYSI-2 scales to the same individual provides another measure of the scale's reliability, in this instance its stability across time, or its test-retest reliability (L. Miller et al., 2011). This measure can provide information about the degree to which an individual's score can change over time. Scales and psychometrics that claim to measure more stable, enduring psychological traits (e.g., personality) or lived experiences should show little change in scores over time. Scales and psychometrics that claim to measure more transient psychological features (e.g., psychological states/emotions) that may vary in response to changes in an individual's circumstances or situations may display change over time. In relation to the MAYSI-2, some scales would be expected to vary at the second testing more than others. These scales include self-reports of emotions such as the items included in the Depressed-Anxious scale. Other scales would be expected to change much less, especially those that are aimed at investigating presence or absence of past behaviour such as the items in the Traumatic Experiences and Alcohol/Drug

Use scales. Test-retest analysis was conducted over a three month and six month delay. The MAYSI-2 test-retest reliability has yet to be assessed over a longer period of time. We expect to see MAYSI-2 forms completed six months apart to correlate less strongly than those completed only three months apart, illustrating the dynamic nature of some of the traits. The rule of thumb used to interpret the correlation coefficients was adopted from Hinkle, Wiersma, and Jurs (2003) and presented in Table 3. Test-retest reliability was conducted using R.

Table 3

The rule of thumb for interpreting the size of a correlation coefficient (r) if n is large

Size of Correlation	Interpretation
.90 to 1.00 (−.90 to −1.00)	Very high positive (negative) correlation
.70 to .90 (−.70 to −.90)	High positive (negative) correlation
.50 to .70 (−.50 to −.70)	Moderate positive (negative) correlation
.30 to .50 (−.30 to −.50)	Low positive (negative) correlation
.00 to .30 (.00 to −.30)	negligible correlation

Descriptive Statistics.

Data were analysed using Statistical Package for Social Sciences (SPSS) Version 23. First, within and between-group statistical analyses were be conducted, comparing MAYSI-2 scores on age, gender, and ethnicity. Descriptive statistics (means and standard deviations) for the sample were also calculated for the SKS. Means and standard deviations of the MAYSI-2 scales are compared to the US norms and other international studies.

Assumptions.

In linear regression the assumptions are on the characteristics of the residuals of the linear regression model. Hence, the assumptions were investigated after performing each linear regression model and are reported in the results section. Because the variables were bounded scales outliers were not a concern (outliers due to data entry errors were checked for prior to analysis). Assumptions for the CFA, and test-retest analysis are also reported in results section.

Chapter 3. Results

Introduction

This study intended to examine the psychometric properties of the MAYSI-2 to determine its usefulness in an NZ context, and provide a picture of the mental health needs of youth offenders in local secure care facilities. Section one of this results chapter presents the findings from the Confirmatory Factor Analysis (CFA) of the MAYSI-2 scale structure, including both evaluation of reliability, convergent, and discriminant validity for the full sample model and for the boys only model. This will be reported after each models factor structure and goodness of fit are presented. The convergent validity of select MAYSI-2 scales will then be presented in section two, looking at the MAYSI-2 ADU scale and the SACS difficulty scale (SACS DS), the MAYSI-2 DA scale and the K6, and the MAYSI-2 SI scale and the Suicide Risk Practice Tool (abbreviated to SS for Suicide Screen). Section three presents a more detailed look at the internal consistency of the MAYSI-2 scales by gender and ethnicity using Cronbach's alpha. The test-retest reliability of the MAYSI-2 scales over a three and six month delay is presented in section four. Section five will summarise the psychometric properties of the MAYSI-2 in NZ, before presenting the last section, which is concerned with the descriptive statistics for the MAYSI-2, SACS, K6, and SS.

The following MAYSI-2 scale abbreviations apply for the results section; ADU = Alcohol/Drug USE; SI = Suicide Ideation; AI = Angry-Irritable; TE = Traumatic Experiences; DA = Depressed/Anxious; SC = Somatic Complaints; TD = Thought Disturbance.

Construct Validity/Confirmatory Factor Analysis

Confirmatory factor analyses were performed to test the factor structure of the MAYSI-2 on an NZ sample using AMOS (Arbuckle & Wothke, 1999). Model fit was assessed using χ^2 value and associated degrees of freedom, Root Mean Square Error of Approximation (RMSEA), and Comparative Fit Index (CFI; Hair et al., 2010). With regards to the χ^2, a good model fit is indicated by the ratio χ^2 / df being as small as possible. A ratio of χ^2 / df $\leq$ 2 is considered a good fit, and χ^2 / df $\leq$ 3 is considered an acceptable fit (Schermelleh-Engel, Moosbrugger, & Müller, 2003). RMSEA values $\leq$ .05 can be considered as a good or "close" fit, values between .05 and .08 adequate, .08 to .10 as mediocre, and > .10 are not acceptable (Browne, Cudeck, Bollen, & Long, 1993; Schermelleh-Engel et al., 2003). The CFI ranges from zero to one with higher values indicating a better fitting model. For this index, a score > 0.9 or close to 0.95 indicates good fit, though ideally it should be closer to 0.95 (Hu & Bentler, 1999; Newsom, 2012).

Empirically based recommendations for sample size were followed prior to conducting the CFA. While a large set of heuristics are available to guide in the selection of sample size, the heavily cited approach indicated by Comrey and Lee (2013), and the stringent 'Rule of 10', first advocated by Nunally and Bernstein (1978) were used. Comrey and Lee (2013) put forward the following recommendations in regards to sample sizes in SEM: 100 = fair, 200 = good, 500 = very good, and >1,000 = excellent. The rule of 10 states that there should be at least 10 cases for each item in the instrument or psychometric being examined. The MAYSI-2

has 52 items and our final sample for CFA was 527. Thus, our sample size satisfied both requirements for conducting CFA with confidence.

CFA were conducted on the full sample, and with boys and girls separately and will be presented in this order. Reporting of the CFA closely follows the recommendations laid out in (Jackson, Gillaspy Jr, & Purc-Stephenson, 2009).

For both the CFA for the full sample, and the CFA conducted for boys, missing data points were present (n=41). Of the 41 missing points, 36 of them were on items which were not used in any of the CFA procedures. Of the five that represented missing data points included in the CFA, missing data imputation was conducted to deal with missing values. The expectation-maximization (EM) algorithm was used, using the missing value analysis procedure in SPSS for single imputation. We did not use multiple imputation because there were only five data points. Age, items that were not analysed (i.e., don't load onto any factors), SACS DS, SACS TU, K6, and SS were all used in the calculation of the imputation in order to enhance accuracy.

Maximum likelihood (ML) estimation method was used to estimate the parameters. The items were not normal due to being binary, but the CFA was conducted as it represented the best option for testing the factor structure and has been shown to be fairly robust to violations to distributional assumptions (Benson & Fleishman, 1994; Savalei, 2008). In cases of non-normality, it is advised to put more weight onto the RMSEA and CFI because using the χ^2 as a measure of model fit will likely be biased towards rejecting true models due to being inflated (Schermelleh-Engel et al., 2003). Non-normality generally does not affect the consistency of parameter estimates (i.e., factor loadings), though one cannot rule out the potential for bias (Bollen, 1989; Savalei, 2008). Parameter standard errors were corrected for non-normality using bootstrap method.

Asymptotically Distribution Free (ADF) method (Browne, 1984) was not used to estimate the parameters even though it has been shown to work well with non-normal (Jones &

Waller, 2015) for two reasons: lack of available statistical packages that allow ADF method and because it has been noted to be biased and unreliable with sample sizes less than 1,000, even for small models (Flora & Curran, 2004).

For each factor, the item with the largest loading was set to 1.0. It is necessary to set one of the loadings to run the model, and choosing the largest one ensures that all loadings are less than 1.0. Due to small sample sizes in ethnic subgroups, CFA was not able to be conducted by ethnicity.

MAYSI-2 CFA model for full sample (n=527).

The model specified that the 37 items (excluding those items that are part of gender specific scales such as TD, TE (boys), and TE (girls)) load on the five latent constructs which represent the MAYSI-2 scales (ADU, AI, DA, SC, SI).

The five factor model indicated a good fit according to the χ^2 / df ratio (χ^2/df = 1009.565/548 = 1.84), the CFI (CFI = 0.909), and the RMSEA (RMSEA = 0.04, 90% CI: .036, .044).

The factor loadings (standardized and unstandardized) of the model are presented in Appendix G, Table 1. Latent variable covariances and latent variable correlations are also presented in Appendix G, Table 2 and 3, respectively.

Only one of the five latent variables showed evidence of convergent validity at the construct level using the AVE method. That scale was SI, which had an AVE greater than .5. The four other MAYSI-2 scales did not show evidence of convergent validity, and showed a high level of variance due to measurement error (refer to Table 4). There also appears to be quite strong correlations between some MAYSI-2 scales such DA and SC (.783), AI and SC (.715), and DA and AI (.666) which are noteworthy. The presence of scale correlations has been noted by the MAYSI-2 authors, who have proposed a plausible rationale (see Grisso &

Barnum, 2006). In short, it is expected that some traits overlap to a degree, as mental and emotional issues are not always discrete phenomena and are commonly comorbid.

In regards to convergent validity at the item level, the standardized factor loadings (pattern coefficients; Appendix G, Table 1) on their designated factors were all statistically significant at $p \leq .002$, with the exception of item 35 (for DA; $p = .035$), and item 47 (for DA; $p = .015$). Strength of standardized factor loadings were mixed, that is to say, 14 of the 37 items were less than the .5 criteria (Hair et al., 2010), and only six of the item loadings on their respective factor were greater than the desired .7 indicating issues with convergent validity.

Table 4

AVE, CR, squared inter-factor correlations, and square root of AVE for the total sample (n=527)

	AVE	CR	DA	ADU	SC	SI	AI
DA	.176	.589	**.419**				
ADU	.342	.804	.286	**.585**			
SC	.308	.724	.783	.421	**.555**		
SI	.575	.868	.577	.180	.466	**.758**	
AI	.320	.801	.666	.476	.715	.523	**.566**

Diagonal items in bold represent the square root of AVE's, and off diagonal items represent squared correlations between scales.

In regards to construct level discriminant validity, only the ADU and SI scales exhibited adequate discriminant validity, as evidenced by the square root of AVE being larger than any of

the inter-factor correlations (refer to Table 4, numbers in bold). SC, DA, and AI failed to provide evidence of discriminant validity, which could reflect items correlating more highly with factors outside their specified factor than with items within their specified factor.

The investigation of structure coefficients was undertaken to assess item level discriminant validity (Appendix G, Table 4). Thompson (1997) and Graham, Guthrie, and Thompson (2003) argue that when the model in question has correlated factors, which our model has, both pattern and structure coefficients should be examined as a failure to do so could lead to interpretation errors. Structure coefficients are used to estimate the relationships between each specific item and each factor. If items in a model are related to factors, they should display a non-zero relationship between them even though the items pattern coefficient (standardized factor loading) has been set to zero (i.e., because the item loading pattern is specified in CFA). This means that while items cannot have a direct effect on factors through loading as in EFA, it may have an indirect effect on the item through its relationship (correlation coefficient) with the factor. Visual inspection of the structure coefficients (see in Appendix G, Table 4) shows that the items on both the DA and AI scale correlate moderately with the SC scale, and thus are reflected in the scale correlations (i.e., they don’t show good discriminant validity). For some items (refer to Appendix G, Table 4; Item13, Item35, and Item39 on the SC scale, Item2, Item11, Item16, Item18, Item 22, Item27, Item28, and Item29 on the DA scale, Item11, Item16, Item17, Item18, Item22, Item28, Item29, Item31, and Item47 on the AI scale), they were empirically indistinguishable from those items that were specified to have a relationship with their parent-factor, in that the relationships were stronger than some of the specified items that were designed to load onto the factor. In other words, the mentioned items loaded onto factors stronger than some items which were designed to. What the above findings indicate is that out of the five MAYSI-2 scales relevant to both boys and girls, only

two show adequate discriminant validity at both the construct and item level (i.e., ADU and SI) in this data set.

Results regarding composite reliability for the five core MAYSI-2 scales are displayed in Table 5 with reference to Cronbach's alpha for the same scale.

Table 5

CR and Cronbach's alphas showing reliability estimates for each MAYSI-2 scale for total sample

	CR	Alphas
DA	.589 [a]	.800
ADU	.804	.787
SC	.724	.694 [a]
SI	.868	.717
AI	.801	.877

[a] ≤ 0.69

We can conclude that the reliability estimated by both Cronbach's alpha and CR is sufficient for the ADU, SC, SI, and AI (if using the more relaxed .6 criteria for Cronbach's alpha generally reserved for exploratory research), however there seems to be no consensus on the reliability of the DA scale. Given CR is generally taken as a superior measure it is more likely the DA scale shows inadequate reliability.

MAYSI-2 CFA model for boys only sample (n=432).

The model specified that the 47 items load on the seven latent constructs which represent the MAYSI-2 scales (ADU, AI, DA, SC, SI, TD, TE (boys)).

The model indicated a good fit according to the χ^2 / df ratio (χ^2/df = 1468.36/878 = 1.67), and the RMSEA (RMSEA = 0.039, 90% CI: .036, .043), though the CFI was lower than desired (CFI = 0.875). The CFI is reported to be highly sensitive to the effects of low average factor loadings, such as those found in general social science fields (Heene, Hilbert, Draxler, Ziegler, & Bühner, 2011). This may to some extent explain the current finding as most factor loadings were indeed low, as will be discussed in relation to item level convergent validity.

The factor loadings (standardized and unstandardized) of the model are presented in Appendix G, Table 5. Latent variable covariances and latent variable correlations are also presented in Appendix G, Tables 6 and 7, respectively.

Similar to the results from the full sample, only the SI scale had an AVE greater than .50. All six other MAYSI-2 scales did not show sufficient evidence of convergent validity and showed a high level of variance due to measurement error. There also appeared to be a number of fairly strong correlations between some MAYSI-2 scales such DA and SC (.831), AI and SC (.682), and AI and DA (.680) indicating the structure does not differentiate the two factors well.

The factor loadings (Appendix G, Table 5) were all significant at the $p \leq .005$ level with the exception of item 47 (for DA; p = .142), and item 35 (for DA; p = .142) which were both statistically non-significant. A large number of factors (21 out of 47) were less than the lower limit of .50; items 8, 7, 47 (on both DA and SI scales), 41, 21, 14, 51 (on both DA and TE), 44, 43, 30, 3, 34, 42, 17, 19 25, 32, 46, and 51. The rest of the factors were greater than .50, and five of the item loadings were greater than the desired .7. The non-significant loadings, and the large number of items that were less than the .5 threshold indicates a lack of convergent validity.

Table 6

AVE, CR, squared inter-factor correlations, and square root of AVE for the boys only sample (n=432)

	AVE	CR	ADU	SC	SI	DA	TE	TD	AI
ADU	.331	.796	**.576**						
SC	.295	.711	.420	**.543**					
SI	.552	.857	.168	.466	**.743**				
DA	.128	.488	.235	.831	.536	**.358**			
TE	.279	.643	.500	.566	.303	.565	**.528**		
TD	.276	.632	.341	.606	.419	.610	.449	**.525**	
AI	.307	.790	.469	.682	.516	.680	.534	.544	**.554**

Diagonal items in bold represent the square root of AVE's, and off diagonal items represent squared correlations between scales.

Very few MAYSI-2 scales showed evidence of discriminant validity at the construct level. Only the ADU and SI scales demonstrated this adequately, as evidenced by the square root of the AVE being larger than any of the inter-factor correlations (refer to Table 6, numbers in bold). The SC, DA, AI, TD, and TE scales all failed to demonstrate adequate discriminant validity using the Fornell and Larcker (1981) method. Investigation of the structure coefficients (Appendix G, Table 8) revealed moderate relationships between most items on each parent factor and the unspecified factor in question, which explain the inter-factor correlations and indicate a lack of adequate discriminant validity. Similar to the first CFA model with the full sample, some items were empirically indistinguishable from those items that were specified to

have a relationship with their parent-factor, in that the relationships were stronger. What the above findings indicate is that out of the seven MAYSI-2 scales relevant to boys, only two show adequate discriminant validity at both the construct and item level (ADU and SI).

Results regarding composite reliability for the MAYSI-2 boys only sample are displayed in Table 7 with reference to Cronbach's alpha for the same scale.

Table 7

CR and Cronbach's alphas showing reliability estimates for each MAYSI-2 scale for boys

	CR	Alpha
ADU	.796	.792
SC	.711	.774
SI	.857	.649 [a]
DA	.488 [a]	.703
TE	.643 [a]	.863
TD	.632 [a]	.621 [a]
AI	.790	.664 [a]

[a] ≤ 0.69

We may conclude that the reliability estimated by both Cronbach's alpha and CR is sufficient for the ADU, SC, SI, and AI (if using the more relaxed .6 criteria for Cronbach's alpha generally reserved for exploratory research). There seems to be no consensus on the reliability of the DA or TE scales, given conflicting results from both estimate methods, and reliability was insufficient for the TD scale. Given the CR is a more robust and less biased estimate of reliability, we can assume the DA, TE, and TD scales do not display adequate reliability.

MAYSI-2 CFA model for girls only sample (n=95).

The model specified that the 42 items load on the six latent constructs which represent the MAYSI-2 scales (ADU, AI, DA, SC, SI, TE; girls).

The model indicated a good fit only according to the χ^2 / df ratio (χ^2/df = 998.825/684 = 1.46). The RMSEA indicated the fit was adequate (RMSEA = 0.070, 90% CI: .060, .079) and the model fit was poor according to the CFI (CFI = 0.781). It should be noted that the sample size is small (n=95) and the model is large, therefore, no firm conclusions can be made about the factor structure for girls. In light of this, convergent and discriminant validity were not calculated for this model. Due to small sample sizes in ethnic subgroups, CFA was not able to be conducted by ethnicity.

Convergent Validity

The MAYSI-2 Alcohol/Drug Use scale and the SACS difficulties scale.

The relationship between the scores on the MAYSI-2 ADU scale and the SACS difficulty scale (DS) was examined using linear regression, with the MAYSI-2 ADU scale as the dependent variable and the SACS DS as the independent variable. Preliminary analysis was conducted by fitting a linear line to look at the relationship between the two variables. Plotting the residuals showed concerns with the normality assumption. The linear correlation between ADU and the SACS DS =.524 ($p<.0001$; R-squared = .274), but plotting revealed a curved relationship (using a lowess line) between the scales (refer to Figure 3).

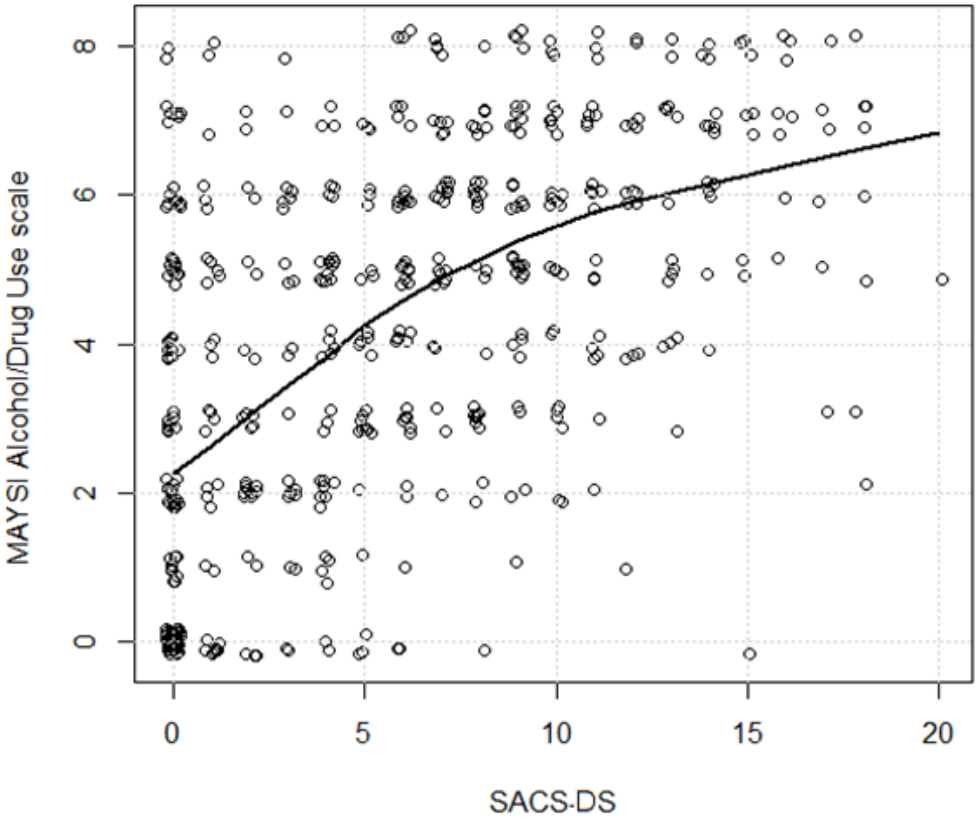


Figure 3. Scatter plot illustrating the non-linear relationship between the MAYSI-2 ADU scale and the SACS difficulty scale (jitter was used to view overprinting).

Given the relationship did not appear to be linear, a polynomial (quadratic/squared) term was added to the regression model. This gave an R-squared of .295. The model equation was 2.395 + .458*SACS.DS - .014*(SACS.DS^2).

Adding the squared term provided a statistical model that was very close to the observed relationship (refer to Appendix G, Figure 1). Having this statistical model then allowed the ability to quantify the significance of the overall relationship between SACS DS and ADU scale which was found to be highly significant (F(df=2, 521)=109, p<.0001). Adding

a quadratic term also meant that there were no longer concerns with the assumptions of normality of residuals (refer to Appendix G, Figure 2). In addition, the residual versus fitted value plot showed no concerns with homoscedasticity (refer to Appendix G, Figure 3). There were no linearity outliers because the data was bounded.

Table 8

Regression output from quadratic equation model for MAYSI-2 ADU scale and SACS difficulty scale

	Beta coefficients	Std. Error	t value	p-value
(Intercept)	2.397	0.16502	14.528	< .0001
SACS.DS	0.455	0.05330	8.528	< .0001
I(SACS.DS^2)	-0.014	0.00349	-3.911	< .0001

N=524; Residual standard error: 2.071; R-squared: .295; F-statistic: 109 on 2 and 521 DF, p-value: < .0001

The above results indicate an overall significant positive relationship between the MAYSI-2 ADU scale and the SACS DS, though visual inspection of Figure 3 revealed two interesting findings. First, it appears that scoring a 0 on the SACS DS does not mean the same as scoring a 0 on the MAYSI ADU scale. The strong clustering at the 0 mark for the SACS DS along the MAYSI-2 axis shows that a large portion of youth who score 0 on the SACS DS score higher on the MAYSI-2 ADU. The vice-versa also applies; those who score 0 on the MAYSI-2 ADU appear to rarely score highly on the SACS DS scale. Second, once those who score 0 on the SACS DS are ignored, it appears there is clear relationship between the two variables. However, it also appears the MAYSI-2 ADU and the SACS DS are measuring two

different things, or not enough of the same thing for either tool to be consistent with the other. We did not expect individuals who took these two tests which purportedly measure the same thing to differ by such an extent.

The MAYSI-2 Depressed/Anxious scale and the Kessler Psychological Distress Scale (K6).

The relationship between the scores on the MAYSI-2 DA scale and the K6 was examined using linear regression, with the MAYSI-2 DA scale as the dependent variable and the K6 as the independent variable.

Preliminary analysis was conducted by fitting a linear line to look at the relationship between the two variables. Investigation of output plots revealed concerns with the assumption of normality of residuals (refer to Appendix G, Figure 4). Normality was not improved by adding a quadratic term but the linear model is robust to departures of normality and therefore we continued with the analysis. In addition, the residual versus fitted value plot showed no concerns with homoscedasticity (refer to Appendix G, Figure 5).

It was determined that the relationship was not linear but curved (refer to Figure 4). The linear correlation between MAYSI-2 DA and K6 =.388 ($p<.0001$; R-squared = .150). To overcome the violation of assumption of linearity, a polynomial (quadratic/squared) equation was added which fit the data better. The squared term added to the regression model gave an R-squared of .170. The model equation was 1.074 + .319*K6 - .010*(K6^2).

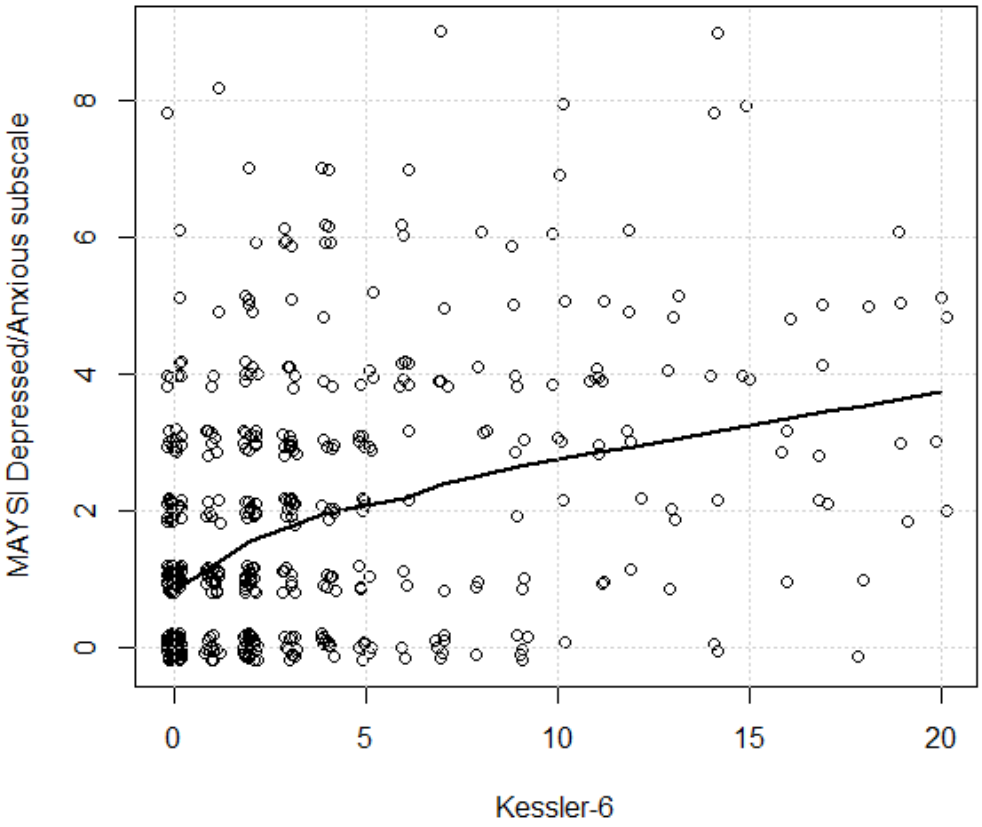


Figure 4. Scatter plot illustrating the non-linear relationship between the MAYSI-2 DA scale and the K6 (jitter was used to view overprinting).

Visual inspection of the plots showed the quadratic model was not as good a fit as the ADU-SACS model, however it is still a small improvement over the linear model and the quadratic term was very significant (F(df=2, 520)=53.3, p<.0001), so this model was used to estimate the relationship between K6 and DA (refer to Appendix G, Figure 6. for visual representation of quadratic model).

Table 9

Regression output from quadratic equation model for MAYSI-2 DA scale and K6

	Beta coefficients	Std. Error	t value	p-value
(Intercept)	1.07	0.11	9.51	< .0001
K.6.TS	0.32	0.05	6.65	< .001
I(K.6.TS^2)	-0.01	0.003	-3.45	< .001

N=524; Residual standard error: 1.746; R-squared: .170; F-statistic: 53.25 on 2 and 521 DF , p-value: < .0001

The above indicates an overall significant relationship between the MAYSI-2 DA scale and the K6, though while significant, the spread of data (refer to Figure 4) invites caution. The heavy clustering at point 0,0 may have influenced the significance of the test, allowing more variation in scores while still being significant. The low R-squared value suggests that the regression model can only explain a small amount of the variation in the data, and that there is a weak relationship between the two variables.

The MAYSI-2 Suicide Ideation scale and the Suicide Screen.

The relationship between the scores on the MAYSI-2 Suicide Ideation scale and the Suicide Screen was examined first using a t-test, and then the non-parametric Wilcoxon Rank Sum Test due to violations in the normality assumption. The t-test was conducted first: difference in means = 2.517 (t-test: t=5.465, df=20.61, p<.0001). The t-test showed significant difference between mean SI for different SS groups which indicated that those with SS small mean SS=1 had larger mean. However, because SI was very skewed and not normal the non-parametric Wilcoxon Rank Sum Test was used. This also found a significant difference in the

distribution of SI between those who scored positive to the SS and those who did not (W = 1941.5, p-value = < .0001). That is, those who scored a 1 on the SS were statistically more likely to score highly on the MAYSI-2 SI scale, and those who scored 0 on the SS were much less likely to score highly on the SI scale. This can be seen when referring to Figure 5 and Table 10. The median SI score for the SS=1 group is far greater than that or the SS=0 group (4 and 0 respectively).

Table 10

Mean, median, standard deviation, and interquartile range (IQR) of the SI compared to the SS score of 0 or 1

	SS=0 (n=502)	SS=1 (n=21)
Mean SI	0.578	3.095
Median of SI	0	4
SD of SI	1.261	2.095
IQR	1	5

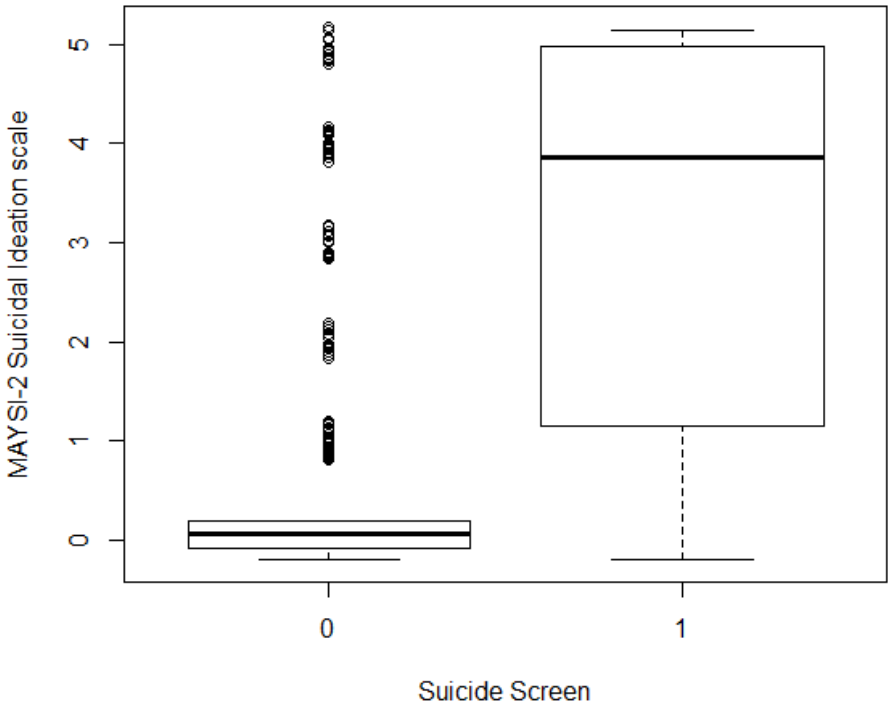


Figure 5. Box plot showing spread of scores on SS in relation to scores on the MAYSI-2 SI scale.

Internal Consistency

The results with respect to reliability are presented in Table 11 which gives alpha coefficients for the seven scales included in the MAYSI-2. For the entire sample, the internal consistency scores (α) for the MAYSI-2 scales ranged from .621 for TD (boys), to .877 for SI. For the most part, MAYSI-2 scales for the total sample were above .70.

In regards to the ADU scale, all groups besides Pacifika girls (α = .687) had α coefficients > .70, with most groups being above .80.

Alpha coefficients for the AI scale were all above .70 with the exception of NZ European boys (α = .670).

The internal consistency of the DA varied between groups. For the whole sample the Cronbach's alphas were all equal to or less than .69, with the exception of the Māori sample (α = .718). The DA scale was found to be much less reliable for boys regardless of ethnicity compared to girls of any ethnicity.

In regards to the SC scale, both NZ European boys and Pacifika boys had α coefficients less than .69 (α = .590; α = .656), while all other scales had reliability estimates greater than .70.

The alpha coefficients for the SI scale were consistently greater than .70, with most being greater than .80. The only exception for this scale was for Pacifika boys, which the scale was not as reliable for (α = .656).

Alpha coefficients for the TD (boys) scale were all less than .69 suggesting the scale is not very reliable, regardless of ethnicity.

The MAYSI-2 TE (boys) scale was similar to the TD (boys) scale, in that all alpha coefficients were less than .69, with two being less than .59 (for NZ Europeans and Pacifika). This suggests the TE scale for boys is not reliable for NZ youth. The TE (girls) scale was more reliable, with an alpha coefficient greater than .70 for the overall sample (all girls). However, for both Māori and Pacifika girls the scale was less reliable (α = .688, α = .680, respectively).

Table 11

Cronbach's Alpha coefficients for each of the MAYSI-2 scales by total sample, gender, and ethnicity

		Cronbach's Alpha (α)			
	Number of items	*Total Sample*	*Māori*	*NZ European*	*Pacifika*
All (both boys and girls)		(n=527)	(n=402)	(n=67)	(n=55)
Alcohol/Drug Use	8	.800	.791	.842	.806
Angry-Irritable	9	.787	.793	.786	.786
Depressed-Anxious	9	.694 [a]	.718	.584 [aa]	.621 [a]
Somatic Complaints	6	.717	.732	.629 [a]	.689 [a]
Suicide Ideation	5	.877	.875	.895	.786
Boys		(n=432)	(n=327)	(n=58)	(n=45)
Alcohol/Drug Use	8	.792	.779	.825	.820
Angry-Irritable	9	.774	.783	.670 [a]	.764
Depressed-Anxious	9	.649 [a]	.680 [a]	.551 [aa]	.506 [aa]
Somatic Complaints	6	.703	.724	.590 [aa]	.656 [a]
Suicide Ideation	5	.863	.853	.902	.656 [a]
Thought Disturbance (boys)	5	.621 [a]	.640 [a]	.579 [aa]	.522 [aa]
Traumatic Experiences (boys)	5	.664 [a]	.670 [a]	.664 [a]	.604 [a]
Girls		(n=95)	(n=75)	(n=9)	(n=10)
Alcohol/Drug Use	8	.833	.838	.935	.687 [a]
Angry-Irritable	9	.809	.797	.829	.844
Depressed-Anxious	9	.781	.786	.750	.759
Somatic Complaints	6	.737	.730	.839	.725
Suicide Ideation	5	.902	.906	.862	.894
Traumatic Experiences (girls)	5	.705	.688 [a]	.777	.680 [a]

[a] ≤ 0.60-0.69, [aa] ≤ 0.50-0.59

Test-Retest Reliability

The test-retest reliability of the MAYSI-2 was estimated over two time frames; three months and six months, by way of Pearson's product-moment correlation coefficient, which is the recommended method for this sort of procedure (L. Miller et al., 2011). Results are displayed below in Table 12.

Table 12

Pearson's r coefficients, 95% confidence intervals and sample size for three month and six month delays

	3 month delay			6 month delay		
MAYSI-2 scale	r	95% CI	n (pairs)	r	95% CI	n (pairs)
ADU	.575**	.213, .798	23	.611***	.479, .716	111
AI	.840***	.655, .930	23	.543***	.397, .663	111
DA	.850***	.673, .935	23	.594***	.459, .703	111
SC	.380 ns	-.038, .685	23	.424***	.258, .567	110
SI	.722***	.441, .874	23	.391***	.220, .538	111
TD	.013 ns	-.444, .465	19	.441***	.255, .595	88
TE	.630***	.294, .827	23	.560***	.417, .676	111

*** $p \leq .001$, ** $p \leq .01$, * $p \leq .05$, ns $p > .05$

With reference to the above table, most MAYSI-2 scales were more strongly correlated over the shorter time period (three months) than the longer delay (six month), with the exception of the ADU scale. However, for this scale, both coefficients were close and the 95% CI overlap substantially. Statistically non-significant results for the SC and TD scales over the three month delay may be partially due to small sample size, though could equally reflect a lack

of reliability over time. This could in theory be expected given they are often transient states. All Pearson's correlation coefficients indicated positive correlations, and the ADU, AI, DA, SI, and TE scales all showed moderate to high correlations over the three month delay (refer to Table 3 for interpretations of the size of a correlation coefficient). As expected, Pearson's correlation coefficients for the six month delay varied from low (SI, TD, SC), to moderate (ADU, AI, DA, TE) for the MAYSI-2 scales. Interestingly, the TE scale, which was hypothesised as being the most enduring, has a smaller correlation coefficient than a number of the other scales which commonly represent more dynamic states (i.e., DA, SI). Overall, the MAYSI-2 demonstrated satisfactory test-retest reliability over a three month period, though by the six month mark, the reliability as represented by Pearson's correlation coefficient, was noticeably poorer.

Summary of MAYSI-2 Psychometric Properties

The CFA model for the full sample (n=527) indicated a good fit according to all measures and indexes. The CFA model for the boys only sample (n=432) indicated a good fit by two measures, but not the third (CFI). Hair et al. (2010) note that CFI values less than .9 are not usually found in models that fit well.

Convergent validity was established for the SI scale in the full sample and boys only model; all other MAYSI-2 scales did not show adequate convergent validity. Item loadings were lower than desired for both models, though more were below the cut-off in the boys only model. There was also fairly strong inter-factor correlations, suggesting latent variables were not measuring discrete traits and might benefit from a second-order factor.

Discriminant validity was established for the ADU and SI scales in the full sample and boys only model; all other MASYI-2 scales did not show adequate discriminant validity.

Composite reliability as assessed by the CFA and internal consistency as assessed by Cronbach's alpha indicated that not all scales used in the MAYSI-2 are consistent in their measurements and are thus limited in reliability. For the full sample and boys only models, the ADU, SC, SI, and AI scales displayed sufficient discriminant validity as evidenced by their CR value. Cronbach's alpha values were mixed, but mirrored the CR values for the full sample and boys only. Reliability estimates for the MAYSI-2 scales with NZ European boys and Pacifika boys appeared to lowest, and the TD (boys) and TE (boys) scales were consistently below a desired level, regardless of ethnicity.

The two scales that satisfied convergent and discriminant validity in both the CFA models (ADU, SI) failed to adequately satisfy convergent validity when compared to psychometric scales previously validated in NZ. The R squared values were much smaller than desired, indicating relatively weak relationships between scales that should, in theory, be measuring the same construct.

Test-retest reliability of the MAYSI-2 scales over the three month time period showed stronger correlations over the six month period with the exception of the ADU scale in which scores from the same individual were correlated more highly, though only slightly. The scores on the ADU, AI, DA, SI, and TE scales showed moderate to high positive correlations over the three month period indicating good test-retest reliability. The SC and TD scales were statistically non-significant and therefore cannot be interpreted. Over the six month delay period correlations ranged from low (e.g., SC, SI, and TD), to moderate positive (ADU, AI, DA, and TE), indicating mixed test-retest reliability over the longer time period.

Descriptive Statistics

In light of the psychometric properties of the MAYSI-2 not holding for NZ youth, the results of the MAYSI-2 scale scores should be viewed with caution. Interpretation of the scores

will be minimal given no firm conclusions can be made as to whether youth indeed exhibit or suffer from the behavioural and emotional problems assessed by the MAYSI-2.

Results of the current MAYSI-2 scores in relation to international studies are presented in Appendix H, Table 1.

Table 13

Mean scores (and standard deviations) on the MAYSI-2 for ethnicity and gender

		Ethnicity		
	Cut-offs	Māori (n=404)	Pacifika (n=55)	NZ European (n=67)
Alcohol/Drug Use	4 of 8	4.32 (2.42)	3.71 (2.58)	4.34 (2.63)
Angry-Irritable	5 of 9	3.55 (2.67)	2.89 (2.55)	4.31 (2.42)
Depressed-Anxious	3 of 9	1.86 (1.97)	1.62 (1.71)	1.90 (1.72)
Somatic Complaints	3 of 6	1.78 (1.77)	1.55 (1.64)	2.01 (1.71)
Suicide Ideation	2 of 5	0.63 (1.34)	0.40 (0.99)	1.18 (1.79)
Thought Disturbance*	1 of 4	0.49 (0.94)	0.40 (0.75)	0.63 (0.99)
Traumatic Experiences	3 of 4	1.98 (1.57)	1.75 (1.54)	2.00 (1.57)

*For boys only

		Gender	
	Cut-offs	Male (n=434)	Female (n=95)
Alcohol/Drug Use	4 of 8	4.23 (2.43)	4.39 (2.61)
Angry-Irritable	5 of 9	3.39 (2.58)	4.59 (2.79)
Depressed-Anxious	3 of 9	1.70 (1.77)	2.53 (2.39)
Somatic Complaints	3 of 6	1.66 (1.69)	2.39 (1.90)
Suicide Ideation	2 of 5	0.58 (1.27)	1.20 (1.83)
Thought Disturbance*	1 of 4	0.50 (0.92)	NA
Traumatic Experiences	3 of 4	1.95 (1.56)	2.06 (1.61)

*For boys only

Table 14

Percentage rates of Scoring Levels across the MAYSI-2 Items

MAYSI-2 Scales	Scoring Level	Sample n=529	
		n	%
Alcohol/Drug Use	Normal	197	37.2
	Caution	135	25.6
	Warning	197	37.2
Angry-Irritable	Normal	333	62.9
	Caution	142	26.9
	Warning	54	10.2
Depressed-Anxious	Normal	357	67.5
	Caution	142	26.9
	Warning	30	5.7
Somatic Complaints	Normal	349	66.1
(n=528)	Caution	166	31.4
	Warning	13	2.5
Suicidal Ideation	Normal	439	83
	Caution	22	4.2
	Warning	68	12.9

Thought Disturbance	Normal	305	70.4
(males only; n=433)	Caution	70	16.2
	Warning	58	13.4

Note: one individual was removed from analysis which used the TD scale and one for the SC scale, as they had missed a number of questions that made up the scale, in turn making it invalid as per MAYSI-2 manual guidelines (Grisso & Barnum, 2006).

What can be said about the findings in relation to the MAYSI-2 scores is that regardless of the specificity of the MAYSI-2 to detect discrete psychopathology, this sample demonstrates high levels of mental health difficulties.

While transferring the scores from the paper MAYSI-2 onto an excel spreadsheet it became apparent that a number of papers were scored incorrectly. This resulted in the author remarking all MAYSI-2 papers to ensure a correct calculation of scale scores. Unexpectedly, 19% of the MAYSI-2 papers were underscored, and thus scored incorrectly.

Descriptive statistics for the SKS.

The SACS.

Descriptive statistics for the SACS were calculated to better understand the alcohol and drug use patterns of youth offenders in secure care facilities throughout NZ and are presented below in Table 15.

Table 15

Results from the SACS (use over the last month; n=524) ranked by frequency of use

SACS question items	Didn't use	Once a week or less	More than once a week	Most days or more
	% (n)	% (n)	% (n)	% (n)
Tobacco	18.7 (98)	4.4 (23)	5.9 (31)	70.9 (371)
Alcoholic drinks	33.4 (175)	36.5 (191)	16.2 (85)	13.9 (73)
Cannabis	34.9 (183)	17.2 (90)	13.5 (71)	34.5 (71)
Synthetic cannabis	86.5 (453)	6.1 (32)	4.4 (23)	3.1 (16)
Other drugs	91.4 (479)	4.6 (24)	1 (5)	3.1 (16)
Hallucinogens	94.1 (493)	4.8 (25)	0.8 (4)	0.4 (2)
Amphetamines	94.7 (496)	3.4 (18)	1.3 (7)	0.6 (3)
Cocaine	95 (498)	3.6 (19)	0.8 (4)	0.6 (3)
Inhalants	95.6 (501)	2.9 (15)	0 (0)	0.2 (1)
E + party drugs	96.9 (508)	2.9 (15)	0 (0)	0.2 (1)
Sedatives	96.9 (508)	1.7 (9)	0.6 (3)	0.8 (4)
Opiates	98.1 (514)	1.7 (9)	0.2 (1)	0 (0)

It appears that the most common form of substance use by youth offenders before being placed in secure care facilities is tobacco, cannabis, and alcohol. All other substances were reported to be used far less. Roughly 35% of all youth in the sample use cannabis most days or more over the month leading up to their placement in a youth justice facility, and roughly 65% had used in the last month. Anecdotally, the 'Other drug' category was predominantly made up of herbal highs and synthetic cannabis, as the data entry category for 'Synthetic cannabis' on CYRAS was incorrectly labelled 'BZP' (a stimulant drug often used recreationally for euphoric

experiences). Thus, both the 'Synthetic cannabis' and the 'Other drugs' categories are majority made up of herbal highs and synthetic cannabis. When taken together, the rates of synthetic and herbal high use become more concerning. Roughly five percent of youth reported having used hallucinogens, amphetamines, cocaine, inhalants, party drugs/ecstasy, and sedatives within the last month, though much smaller numbers were using these drugs regularly.

Table 16

Means and standard deviations for the SACS DS, Kessler 6 and Suicide Risk Practice Tool by gender and ethnicity

	SACS DS	Kessler 6	Suicide Risk Practice Tool[a]
	Mean (SD)	Mean (SD)	% (n)
Full sample (n=524)	5.88 (4.99)	3.49 (4.52)	4 (21)
Māori (n=399)	6.08 (4.95)	3.50 (4.61)	4 (16)
Boys (n=325)	*6.27 (4.98)*	*3.31 (4.42)*	*3.4 (11)*
Girls (n=73)	*5.23 (4.77)*	*4.33 (5.32)*	*6.8 (5)*
Pacifika (n=55)	4.29 (4.73)	2.71 (3.57)	5.5 (3)
Boys (n=45)	*4.42 (4.74)*	*2.56 (3.62)*	*4.4 (2)*
Girls (n=10)	*3.70 (4.90)*	*3.40 (3.44)*	*10 (1)*
NZ European (67)	5.93 (5.20)	4.16 (4.68)	3 (2)
Boys (n=58)	*6.31 (5.23)*	*3.71 (4.12)*	*3.4 (2)*
Girls (n=9)	*3.44 (4.45)*	*7.11 (6.98)*	*0 (0)*
Other (n=3)	7.67 (7.01)	1.67 (2.89)	0 (0)

[a] Percentage and number represents those who were deemed a risk by the suicide risk practice tool.

Overall, scores on the SACS DS indicate that on average, youth in secure care facilities in NZ display substance use problems which likely require intervention (refer to Table 16). Further, all the means were greater than 2; a score of 2 or above indicates a need for further assessment. To be specific, 71.9% of the sample score 2 or above, indicating over two thirds of the sample were in need of further assessment. Sixty-three percent of the sample scored 4 or higher and 50% scored a 6 or higher. Pacifika youth of both genders and NZ European girls scored on average in a range considered to be of clinical significance, and that likely requires intervention (≥4). The means of Māori and boys that were Pacifika or NZ European show that youth belonging to these subgroups on average scored in the highest range (≥6) indicating serious problems which may require intervention from a specialist substance use service. Interpretation of the 'Other' category is made ambiguous by the small sample size, and the scores presented may not be a true representation of this groups substance use needs. Broad gender differences seem evident, with girls on average scoring lower than boys, regardless of ethnicity. This finding is interesting as it differs from the MAYSI-2 ADU scale, in which girls scored slightly higher on average than boys (refer to Table 13).

The K6.

In regards to the scores on the K6, scores for the full sample were on average within the range indicating that they are most likely well or absent of mental illness in the previous 30 days (i.e., between 0 and 7). Only NZ European females were within the 8-12 range. That indicates they are likely to have a mild to moderate mental health disorder. This finding is important because it seems to suggest that on average, these youth do not display symptoms of psychological distress that warrant concern, which, given the plethora of literature on the topic, seems highly unlikely. In light of this finding percentages were calculated for each subgroup to

allow comparisons with the results of a study by Krynen et al. (2013). That study looked at the item response properties of the K6 in a large NZ sample (n=4401). These percentages are presented in Table 17. Between 80-90% of all youth in the sample were deemed 'mentally well', with slight variability between ethnicities.

Table 17

Percentages for Low, Mild/Moderate, and Severe Non-Specific Psychological Distress (K6) categories for Māori, Pacifika, and NZ European youth offenders in secure care facilities in NZ

	Ethnicity		
K6 category	Māori (n=399)	Pacifika (n=55)	NZ European (n=67)
No psychological distress (score 0-7)	84.5% (n=337)	90.9% (n=50)	79.1% (n=53)
Mild/Moderate psychological distress (score 8-12)	8.0% (n=32)	7.3% (n=4)	14.9% (n=10)
Severe psychological distress (score 13-24)	7.5% (n=30)	1.8% (n=1)	6.0% (n=4)

The suicide risk practice tool.

Findings from the suicide risk practice tool need to be interpreted in accordance with the sample sizes of each subgroup. For example, 10% of PI girls were deemed at risk for suicide, though the subsample only consisted of 10 individuals; and 33% of the 'other'

ethnicity reported being at risk, though there were only 3 in the subgroup. From subgroups with adequate numbers, percentages determined at risk of suicide seemed to be around 4-5%.

Summary of descriptive statistics results.

As previously stated, results from the MAYSI-2 must be interpreted with caution given the psychometric properties displayed with our sample of NZ youth. Briefly, differences between ethnicities were evident, with Māori and NZ European youth scoring higher than Pacifika youth on all MAYSI-2 scales. Gender differences were also present; females scoring higher on all MAYSI-2 scales compared with males. The most highly endorsed scale was the ADU scale, with over 60% reaching either caution or warning cut-offs, and thus needing further attention. AI, DA, and SC scales were all similarly endorsed, with approximately 35% of youth reaching warning or caution cut-offs.

Results from the SACS show our sample of youth offenders in secure care facilities most frequently endorsed the use of tobacco over the previous month, with over 70% using ‘most days or more’. Alcohol was the second most common substance to be endorsed, though most youth only stated they had consumed it ‘once a week or less’ over the past month. Cannabis was endorsed slightly less than alcohol, though the frequency of its use was much greater, with approximately 35% of youth reporting they had used cannabis ‘most days or more’ over the last month. All other substances were endorsed by less than 10% of the sample.

Scores from the SACS DS indicate a high level of substance use among youth offenders in secure care facilities; levels which require further assessment and intervention. Both Māori and Pacifika boys scored in the highest range, though females, and NZ Europeans still on

average, scored in a range which likely requires further investigation. Over two thirds required further assessment and half likely need intervention from specialist services.

At a sample level, the K6 indicated that the average youth was most likely well or absent of mental illness. At a group level, the NZ European females were on average more likely to fall within a range that indicated a probable mild to moderate mental health disorder. NZ European were also more likely than Pacifika youth and Māori youth to have some level of psychological distress.

Suicide risk as detected by the suicide risk practice tool placed risk of suicide for the full sample around 5%, with girls showing slightly elevated risk. Small sample sizes of subgroups limit firm conclusions.

Chapter 4. Discussion

Introduction

The main purpose of this research was to determine the usefulness of the Massachusetts Youth Screening Instrument, Version 2 (MAYSI-2) in providing a preliminary assessment of the mental health needs of NZ youth offenders in secure facilities. It also aimed to provide a better understanding of the mental health needs and characteristics of adolescents in secure care facilities within a NZ context. This chapter seeks to provide an overall discussion and integration of the current findings by summarising the results within the context of the reviewed literature. First, the MAYSI-2 findings will be discussed in relation to international findings. Possible reasons for the findings relating to the MAYSI-2 reliability and validity status in our sample of youth offenders will also be put forward. The results from the SACS,

K6, and the Suicide Risk Practice Tool will be discussed with reference to both general NZ population and international studies of youth offenders in secure care facilities. Secondly, the implications of the current research and general recommendations will be presented, followed by the strengths and limitations. Lastly, directions for future research will be considered.

The MAYSI-2

The current study found that while the 5 factor model (full sample) of the MAYSI-2 was a good fit, it failed to adequately satisfy reliability and validity conditions, meaning the MAYSI-2 is unlikely to be accurate or consistent within the NZ context. Further, the 7 factor model (boys only) failed on one index of goodness of fit, suggesting the model does not fit appropriately, and similarly did not adequately satisfy reliability and validity conditions. While these results are not encouraging for the continued use of the MAYSI-2 in NZ, the results are consistent with the growing literature indicating problems with the application of the MAYSI-2 in contexts and cultures within and outside the US (Cauffman & MacIntosh, 2006; Colins et al., 2014; Lennox et al., 2014; McCoy, 2010, 2011; Stathis et al., 2008). While it is impossible to narrow down the exact reasons for the MAYSI-2's lack of reliability and validity in the sample used in this study, there are a number of possible explanations.

The first is the cultural linguistic differences between youth in secure care facilities in the US (where the MAYSI-2 was developed) and in NZ. Just under 80% of our sample identified as NZ Māori, and at this stage we are unable to assume with any level of certainty that the language used for the questions in the MAYSI-2 are interpreted the same by Māori (or Pacifika), and African American, Hispanic, and NZ European youth. As highlighted by Kersten et al. (2016), issues may arise due to the content of the questions that make up each scale. Anecdotal reports from Stathis et al. (2008) study found the indigenous youth of Australia would frequently get confused when filling out the MAYSI-2 and struggled to answer a number

of questions. Both the nature of the questions, and the underlying concepts were reported to be misunderstood by a number of the indigenous youth, resulting in underreporting of symptoms. Stathis et al. (2008) concluded that the MAYSI-2 may be better replaced by a tool that is more appropriate for use with Australian indigenous youth. McCoy (2010) further investigated this phenomenon in African American and Caucasian youth in the US using qualitative methodology. McCoy (2010) identified a number of questions that were confusing for African American youth, which ultimately resulted in those youth responding in the following ways: minimizing their responses, withholding information, and responding inaccurately.

Culture provides a lexicon and framework to understand emotions, making some emotions more salient, and others more difficult to identify and describe (Kirmayer, 2001). In light of this knowledge, we are unable to rule out the presence of cross-cultural construct and item bias. That is, constructs measured may not be identical and may have different meanings to the distinct cultural groups present in this study (He & van de Vijver, 2012).

The results may also be explained by a number of other factors. It is possible that Māori and Pacifika individuals performed differently on MAYSI-2 subscales because of differences in perceived normality of the behaviours assessed. The experience of illness and disease is culturally bounded, influencing many aspects, including the perception, interpretation, and expression, of psychological symptoms (Kirmayer, 2001). Differences in willingness to disclose information may have resulted in underreporting of some symptoms, and the constructs measured may not accurately reflect mental health constructs in Māori and Pacifika cultures. Similarly, socio-culturally acceptable patterns of emotional expression could be different to those that the MAYSI-2 was developed to assess for. The above factors may have ultimately led to differences in reporting patterns which resulted in discrepancy between the NZ and US factor structure/models.

Local research supports the possibility that cultural differences may have played a role. A study conducted in NZ using the Symptom Checklist-90-Revised (SCL-90-R), a tool for measuring psychopathology, found cultural identity significantly affected performance on the scales (Barker-Collo, 2003). A recent study also indicated Māori parents filling out the SDQ for their child did not generally support the questionnaire approach and expressed a desire for a more informal, face to face meeting (Kersten et al., 2016). Ethnic minorities (e.g., Māori, Pacifika) also reported difficulties with words in the SDQ questionnaire, finding many of the questions confusing. The authors concluded by calling for future work to develop a tool that is culturally appropriate and valid.

On a similar note, Cauffman and MacIntosh (2006) suggests the ability of the MAYSI-2 to accurately screen mental and behavioural problems in African American youth offenders may be impaired. Further, the psychometric properties of the MAYSI-2 has been questioned with its use on minority populations, as shown in Colins et al. (2014). The reliability and validity of some MAYSI-2 scales were not well supported, and the authors concluded the use of these scales to make clinical decisions was not advised.

Taking the above into consideration, it seems possible that some ethnic groups may have inconsistently reported some symptoms, for a variety of reasons, or the MAYSI-2 psychometric properties may not have held for some ethnicities in the current sample. To what little or large extent, we do not know, though with the available data and in light of the aforementioned studies, we are not in a position to rule out the possibility of culture playing a role in the limited validity and reliability of the MAYSI-2 in the current sample. In light of the above results, the prudent option is to discontinue using the MAYSI-2 to screen for the mental health needs of youth offenders in NZ because it is unlikely to be accurate.

The unexpected finding that just under 20% of the MAYSI-2 forms were scored incorrectly is of serious concern. The MAYSI-2 is used to inform future care, ongoing assessment and monitoring, and possible referrals to specialist services. A number of the youth who were underscored were scored below a cut-off that, if scored correctly, they would have scored above the threshold and therefore had further assessment and possibly been referred to specialist service. This finding highlights that staff administering and scoring the MAYSI-2 need further training and supervision on administration and scoring.

Rates of Mental Health Needs

Given this study is the first conducted in NZ to look broadly at the mental health needs of youth offenders in secure care facilities, it is difficult to place the findings into the local context. The following section will attempt to do so by discussing each key finding in relation to local research on community samples of adolescents as well as international studies focussed on youth in secure care.

Substance Use Problems

Results from the SACS indicated high use of tobacco, alcohol, and cannabis in the current sample, though relatively less usage of other substances. As a discrete population, youth offenders in secure facilities appear to suffer from high rates of substance use disorders – this finding is well replicated and our data suggests the same trend holds in the current sample (Gretton & Clift, 2011; Harzke et al., 2012; Ståhlberg et al., 2010).

Unfortunately, the SACS does not imply or indicate diagnosis of any substance use/misuse disorder, thus limiting direct comparisons to the large majority of studies which use diagnosis as an indicator. Further, despite its wide use in secondary care services throughout NZ, the SACS has not been the subject of much study within NZ and has not yet been trialled

abroad. Comparisons can be drawn with the original SACS study (Christie et al., 2007, 2011) which looked at substance use difficulties in community sample youth and youth who presented at a drug and alcohol clinic in NZ. Youth in the current sample scored disproportionately higher on the SACS than the community sample used in the original study. Youth from secure care facilities were over three times more likely to need further assessment (72% versus 20%), nearly seven times more likely to score in the clinically significant range (63% versus 9%), and ten times more likely to score past the cut-off indicating significant problems which may require intervention from a specialist service (50% versus 5%). When scores from the current sample of youth are compared to the clinical sample who presented at a community alcohol and drug AoD centre we see broad similarities. The majority of youth in both groups required both follow up assessment and likely intervention from services. The scores of the clinical sample were higher than those found in the current study though the rates of identified AoD issues in the current sample are of great concern.

While most studies do not differentiate what substances are being misused when assessing for substance use problems, some of the more thorough state or country wide studies such as those conducted in New South Wales, Australia, do (e.g., Indig et al., 2011). Comparisons with results from this study reveal broad similarities and differences. Far higher rates were found in Australia; for example youth offenders self-reported use of cannabis and alcohol prior to custody compared to the current study. Sixty-four percent reported at least weekly use of cannabis in an Australian sample, compared with 48% who used once a week or more in the current study. With respect to alcohol use, 66% of the Australian sample reported weekly use prior to custody, compared with 30% in the current study. While this finding could represent a real difference between youth offenders AoD use prior to admission, other explanations are also available. The Australian sample included youth offenders up to the age

of 21, whereas the current study was limited to age 17. It is possible that older youth are more likely to be more regular users of AoD, especially alcohol given the ability to purchase it legally after turning 18. This may also explain the larger discrepancy between prior alcohol use between samples. Despite this difference, given the young age of the current sample, the rates of alcohol and marijuana use is of concern.

While the current sample reportedly misused substances at a lower rate than youth offenders in Australia, compared to age equivalent youth from the general population the current sample appear to have much higher rates. Finding direct comparisons is difficult as most surveys conducted in NZ that look at patterns of alcohol consumption in youth are generally looking specifically at problematic drinking such as binge drinking (i.e., quantity), whereas the SACS reports on frequency of alcohol use and other difficulties related to their substance use. However, findings from the Youth 2012 study (students aged 13-18) reported only 8% drank alcohol weekly and only 57% of all students had ever tried alcohol (Adolescent Health Research Group, 2013). Additionally, Ministry of Health data suggests 11.5% of youth aged 15-17 are categorized as “hazardous drinkers” (AUDIT score ≥ 8) and that only 3.4% of youth aged 15-17 consumed 6+ drinks on one occasion at least weekly (Ministry of Health, 2016). Compare this result with current findings; that ~70% of the sample have used alcohol at least once a week and ~30% use more than once a week or most days over the last month. It is clear youth offenders in secure care facilities are a much afflicted population when it comes to alcohol use and misuse.

With regards to cannabis use, findings from the Youth 2012 study indicated 5% of boys and 2% of girls in their sample used weekly or more (3% overall) and only 23% of youth reported that they had ever tried it (Adolescent Health Research Group, 2013). Prevalence for the current sample far exceeds those reported in the community; ~35% reported using most

days or more, almost 15% used more than once a week, and 65% had used at least once in the last month.

Males (of all ethnicities) scored higher than females on the SACS Difficulties Scale, indicating a higher need for clinical intervention. This result is consistent with large scale international studies which indicate a moderate but consistent difference between male and female youth offenders in secure care facilities with a higher prevalence in males with respect to substance use problems (Archer et al., 2010; Domalanta et al., 2003; Lader et al., 2003; Teplin et al., 2002; Timmons-Mitchell et al., 1997).

Considering ethnicity, there was marginal difference between NZ European and Māori youth in how they reported their difficulties with substances on the SACS; Māori youth scored highest and Pacifika youth scored somewhat lower than both. This finding replicates the results found in Christie (2011) in the community sample of youth who took the SACS. Māori youth scored highest, followed by NZ European youth, and then Pacifika youth. It appears the trend of young Māori being disproportionately affected by substance use problems in the community persists into the forensic setting. Larger studies of youth in the community have previously found much larger discrepancies between Māori youth and other ethnicities (e.g., Crengle et al., 2012; Fergusson et al., 1997). The most parsimonious explanation for this finding is that different techniques employed to measure substance use difficulties may exacerbate or reduce this difference between ethnicities. It may also be the case that the pathway to offending for youth in our sample may include substance use which would not be the case for community samples, thus rendering the differences between substance use issues in the current sample lower than in the general population. Pacifika youth scored proportionately less than both Māori and NZ European youth on the SACS DS. This finding is in accordance with studies of community samples, which have found Pacifika students use both alcohol and cannabis less

than NZ Europeans, who in turn use less than Māori (Adolescent Health Research Group, 2013).

Kessler 6

The rates of 'psychological distress' indicated by the K6 screening tool in the current sample are lower than expected. At the sample level, the K6 indicated that the average youth was likely well or absent of mental illness, using the recommended cut-offs. This finding is in direct conflict with the consistent, cross-cultural finding that youth in secure care facilities display high levels of psychopathology (see Appendix A, Table 1). While the K6 looks specifically at depression and anxiety symptoms, this finding is still somewhat surprising given the established rates of depression and anxiety in international studies of youth offenders in secure care appear to be around 20-40% for depression (Domalanta et al., 2003; Hamerlynck et al., 2008; Indig et al., 2011; Pliszka et al., 2000; Teplin et al., 2002) and between 6-20% for anxiety disorders (Archer et al., 2010; Harzke et al., 2012; Teplin et al., 2002; G. Wasserman et al., 2010), and likely much more if those with sub-threshold symptoms were included. While the K6 has not been used specifically within a youth offender population, comparisons can be drawn between general adolescent populations internationally (Chan & Fung, 2014; Green et al., 2010; Mewton et al., 2016; Peiper et al., 2015) and adult populations within NZ (Krynen et al., 2013).

Rates of psychological distress reported in the current study were proportionately lower than results from school aged student samples in Mewton et al. (2016) and Peiper et al. (2015), and substantially lower than reported rates in Chan and Fung (2014). Rates in the current study were more similar to Green et al. (2010) which found approximately 90% of adolescents in their sample reported being 'mentally well'. Keeping in mind that the above literature is concerned with general population samples, the substantially lower rates of reported

psychological distress from our sample is both surprising and somewhat concerning. Compared with K6 results from a large sample of adults in NZ (Krynen et al., 2013), youth in the current sample scored substantially lower on levels of psychological distress. This finding is somewhat concerning given the onset of many mental health difficulties occur in adolescence and thus higher levels would be assumed.

Studies conducted in general population samples of NZ youth fail to find meaningful differences between NZ European and Māori in their reports of depressive symptoms (Crengle et al., 2012) which appears consistent with the current findings; the majority of Māori, NZ European, and Pacifika youth scored in the 'absent of mental illness range'.

Consistent with previous research on the K6, females from all ethnicities reported higher mean levels of psychological distress compared to their male counterparts (Drapeau et al., 2010; Mewton et al., 2016; Peiper et al., 2015). Only NZ European females were, on average, likely to have a mild to moderate mental health disorder. Findings from longitudinal studies indicate females in the general NZ population consistently measure higher in depressive symptoms (Clark, Fleming, Bullen, Crengle, et al., 2013), and international studies of youth in secure care facilities also see a bias in anxiety and mood disorders towards females (for example see Archer et al., 2010).

A number of hypothesis exist for explaining the gender differences commonly seen such as those in the current study (see Drapeau et al. (2010) for a more detailed discussion). One such explanation is the possibility that, of those females that offend, it is only those with the most serious psychopathology and behavioural problems that are court ordered to a secure care facility, given this is more of a punitive response. This hypothesis assumes that the legal system in NZ is more punitive towards young male offenders which may not necessarily be the case. Findings from overseas in adult offender populations have identified males tend to be

sentenced more punitively than females (Spohn & Beichner, 2000; Steffensmeier & Demuth, 2006), so this hypothesis may hold some merit.

Alternatively, it may be that for female youth, offending is a symptom of significant psychopathology, unlike their male counterparts who may engage in illegal behaviour without such levels of psychopathology, due to the impulsive and risk taking behaviours which are common in male adolescents (Byrnes, Miller, & Schafer, 1999). This hypothesis would therefore result in female offenders having substantially higher rates of mental, emotional, and behavioural problems compared to males in secure care facilities. While rates of young female offending has in recent times increased in NZ (Lynch, 2014), there is almost a complete lack of local research on young female offenders (Best, 2013) and the response of the juvenile justice system, which limits our ability to make any meaningful conclusions.

Another possible explanation is that the K6 assesses more for emotional symptoms as opposed to behavioural symptoms of psychological distress. Behavioural symptoms are thought to be more common in adolescents and in males specifically, thus the emphasis on emotional symptoms may not be the best measure because it would result in an inherent response bias (Green et al., 2010). Gender differences in the way females and males perceive and express psychological distress, as well as how comfortable they feel to disclose this may also have impacted on the results. The wording and content of some of the items may be aligned more with how females experience and express their distress. Thus, the sex based differences may be due, at least in part, to question bias. Previous research has looked at including behavioural symptoms to augment the K6 in the hopes of reducing the impact of sex-based measurement non-invariance and enhance the predictive utility with male participants (Green et al., 2010). Findings are promising for the inclusion of behavioural symptoms to augment the K6, though further research is needed.

The rates of reported psychological distress in the current sample begs the question as to whether rates are indeed extremely low, or whether the K6 is not accurately detecting them. Considering the first possibility, given the wealth of literature indicating youth offenders in secure care facilities display significantly higher rates of mental health conditions, including anxiety and depression, it seems unlikely that NZ youth would differ to such a large extent. Further, it seems highly unlikely that youth offenders in secure care facilities would have lower rates than general population youth and adult samples. However, one cannot be certain without conducting further research into the area.

It is possible that due to males being vastly over represented in the current sample (82%) and the possibility of item bias, that the overall scores would be lower. It may also be the case that the K6 is not the most useful tool to use in youth offender populations. While in recent years it has accumulated evidence indicating it as a useful tool to use with adolescent samples internationally (Chan & Fung, 2014; Green et al., 2010; Mewton et al., 2016; Peiper et al., 2015), to the best of the authors' knowledge, it has yet to be established in youth justice populations which are unique in their own right.

Suicide Risk Practice Tool

Overall, 4% of the sample were deemed at risk of suicide by the Suicide Risk Practice Tool. That is, 4% of our sample disclosed that they were thinking about suicide or serious self harm. Compared to general adolescent samples included in longitudinal studies, this appears low. NZ students report much higher rates of suicidal intentions and behaviours, though rates which look specifically at current suicidal ideation appear to be non-existent in community samples of youth within NZ which impedes direct comparison. We can however place these findings in the context of local studies which use 12 month prevalence rates. Within a 12 month period, 21% of female students and 10% of male students experienced suicidal

thoughts, 11.5% of female students and 6.1% of male students made suicide plans, and 6% of female students and 2% of male students had made a suicide attempt (Clark, Fleming, Bullen, Denny, et al., 2013; Fortune, 2010).

International studies of youth in secure care facilities also indicate much higher rates of risk behaviour that would be identified by the Suicide Risk Practice Tool, though this result is likely due to the recall periods being longer than at the present moment (e.g., 'have you in the last few months' compared with 'are you thinking of killing yourself now'). As Stokes et al. (2015) makes clear, prevalence rates of suicide ideation, intentions, and behaviours vary by when the assessment takes place (e.g., intake, during stay, discharge).

Current ideation at intake/admission to a secure facility is the ideal comparison for the current findings, and rates from international studies are very similar to those found in our sample. Prevalence rates of suicidal ideation on admission were 3.0% in Archer et al. (2004), and 5.5% in Penn et al. (2003), closely aligning with the 4% of youth found in the current study.

Minimal difference was found between risk status and ethnicity (refer to Table 16), which appears to contradict findings of studies focussing on general population adolescents. Māori students have been reported to be disproportionately higher than NZ European students with respect to reporting suicidal thoughts and plans (Fortune, 2010). Moreover, Māori students were more likely to self-report that they had made a suicide attempt and deliberately self-harmed (Crengle et al., 2012). Māori youth suicide rates have also been at least 1.7 times the non-Māori youth suicide rates (Ministry of Health, 2015), raising questions as to why a greater disparity is not apparent in our sample of youth offenders in secure care facilities with respects to ethnicity.

A difference between risk status and gender was identified in the current sample with females presenting as at risk about twice as often as males. The ratio of females to males

reporting suicidal ideations is similar to some studies of youth offenders in secure care facilities internationally (Abram et al., 2008; Cauffman, 2004; Vincent et al., 2008), and within community adolescent samples in NZ (2:1, female:male; Clark, Fleming, Bullen, Denny, et al., 2013; Fergusson et al., 2011).

Implications

This study identified current methods for screening young people in secure care facilities may not be providing an accurate and valid representation of their mental health needs. Specifically, the MAYSI-2 should be used with caution and ideally replaced with a measure that displays adequate psychometric properties with youth in NZ. While the MAYSI-2 is the most widely used screening tool in secure facilities internationally, the current study echoes the findings of previous research calling into question its usefulness in novel contexts and with unique populations.

The unexpected finding that almost 20% of the MAYSI-2s administered were scored incorrectly requires further investigation, immediate remediation, and likely further investigation into the consistency of scoring for other measures too. Ideally, ongoing training as to best practice use of psychometrics should be provided for all youth justice staff that are responsible for administration and scoring. At a minimum, specific youth justice staff should be trained in administration and scoring of the relevant psychometrics used. This approach will not only limit the chance of incorrect scoring but also ensure consistency between youth justice facilities.

The picture of the mental health needs assessed in the current sample of youth offenders in secure care facilities within NZ is complex. Rates of substance use were far in excess of community samples indicating a need for specialised treatment. Clinicians and youth justice workers must be aware of the misuse of substances and significant dysfunction associated

within this population. There is now broad consensus on the detrimental effect of regular cannabis and alcohol use on the developing brain (Squeglia, Jacobus, & Tapert, 2009; Volkow, Baler, Compton, & Weiss, 2014), and the levels of use illustrated in the current sample pose a unique challenge to the youth justice system with over half requiring further clinical assessment and likely intervention. Particularly for those who are habitual users or have an undiagnosed substance use disorder, being placed in a secure care facility may lead to withdrawal effects and a potentially increased level of risk. While acknowledging the primary aim of the youth justice system is to reduce recidivism, it is argued that there is an ethical mandate to ensure both physical and mental health needs are addressed sufficiently. Therefore, in addition to identifying the mental health needs of the individual, there also needs to be a clear pathway for treatment. While an assessment of the treatment pathway for youth in secure care facilities is outside the scope of this study, it is important to state explicitly that the need for a clear and well-resourced pathway is essential in order to meet the needs of these at risk youth.

Levels of current suicidal ideation are of concern and, as the leading cause of death in youth justice facilities (McCoy, 2015), a more thorough screen of suicide may be advised, including a history of previous attempts and other clinically relevant risk factors (Lewinsohn, Rohde, & Seeley, 1996). This additional screen may allow for more thorough care planning. Further, due to the acute nature of suicidality, repeated assessment over time should be carried out as it can provide clinically relevant information pertaining to the ongoing care of a youth in a secure facility. The suitability of the K6 measure has been called into question by the current results. It may be the case that youth offenders in secure care facilities display very low levels of ‘psychological distress’ compared to community samples, or it may be a more nuanced approach to assessing for mental health concerns is required.

The goals of reducing re-offending and looking after the welfare of the youth should not be viewed as separate but as complementary. Inherent in looking after the welfare of a youth is making sure their mental health needs are appropriately met. Meeting these needs however, relies on the knowledge of how to best use the resources available to the juvenile justice system in a targeted way. Though, even before this can be done, one must first know what those needs are. Having data on the types and frequency of mental health problems which occur commonly in this population allows progress in a way which satisfies all of the above requisites. While more research is needed to flesh this out further, what is proposed from the data at hand is a system which can respond to acute mental health concerns (e.g., suicidal ideation, alcohol and drug use/withdrawal), as well as more long term interventions which aim to treat the mental and behavioural problems these young people present with. What is hoped is that the boundary between public services and the youth justice system do not interfere with the continuation of mental health treatment after release, and that treatment can be ongoing when required.

Directions for future research

The current research has found that another method of screening this vulnerable population for psychological and behavioural issues when entering secure care facilities in NZ is needed. This conclusion presents itself as an opportune area for future research. One option is to locally develop a psychometrically robust tool to use with the wider population of youth offenders, potentially not just those in secure care facilities. A tool is needed that looks specifically at the common psychological and behavioural issues plaguing this group, and enables youth justice workers to be confident using the results to make decisions about an individual’s care. Given Māori are disproportionately represented in this group, any new tool should be developed with consideration as to how Māori youth experience mental health issues. Cultural consultation throughout the process would be essential. There are a number of other

options available which have relative positives and negatives. A potential option is using another psychometric tool to screen youth offenders in NZ. While alternatives to the MAYSI-2 exist (see Grisso, Vincent, and Seagrave (2005) and Richardson et al. (2015) for reviews), these tools should also pass through the same psychometric testing processes before they are used locally, given they are all developed and normed overseas. The same should apply for those currently in use, including but not limited to the K6.

Specifically with respect to the psychometric validity, future research could investigate the timing of psychometric assessments in the youth justice pathway. It could be the case that completing the MAYSI-2, SACS, K6, and Suicide Risk Practice tool on intake may be overburdening the young person, especially if they have come straight from court or the police cells. Qualitative inquiry may be fruitful in this area, with consultation from youth themselves.

Another suggestion for further research is a more thorough investigation of the mental health needs of youth offenders more generally, and those in secure care facilities. The second aim of this research was impacted by the inability to infer anything from the MAYSI-2 results, and thus the author feels that the question was not answered comprehensively. The likely next step to build a stronger overall picture of the mental health needs of youth offenders is to commission future work building on from what has been established thus far.

A peripheral aim of the current study was to determine whether the needs of this group are being met sufficiently, and future research has the potential to provide a mandate to increase service development and access to specialist services, if indicated. Depending on future findings in the area, future research may focus on the development of specialised treatment programmes that look to target mental and behavioural problems that are common among youth offenders here in NZ (e.g., substance use, depression, anxiety, emotion dysregulation).

Limitations

The findings of this study should be viewed in the context of its methodological limitations. Firstly, it's important to point out the reliance on the participants' self-reporting, which raises the issue or reporting bias. Given the potential anti-social nature of the sample and the sensitive nature of the material discussed, it is possible that underreporting may have occurred due to a reluctance to discuss sensitive and difficult topics (e.g., abuse, suicidal ideation, emotions). It may also be the case that gender differences are, in part at least, due to underreporting of symptoms which may be more prevalent for one gender. Future studies looking at the mental health needs of youth offenders would benefit from cross-informant interviewing to limit the possibility of biased or underreported symptoms.

Data for this study was retrospectively gathered from a non-selected group of youth who were admitted to one of NZ's four youth justice facilities between January 2014 and December 2016. The availability of data was determined primarily by the pro-activeness of the managers in each youth justice facility to deliver the MAYSI-2 forms. As a result, the sample methodology is not as robust and may, in theory, be different than a random sample from the same population. While it is difficult to determine exactly how this sampling method could have influenced the sample composition and subsequent results, it cannot be ruled out as having no effect.

While numbers from the three North Island facilities were roughly similar, substantially lower numbers were gathered from the only South Island facility. While there did not appear to be an obvious difference in gender or ethnic composition (see Table 2), the under inclusion of youth from the South Island facility means the sample may not be as precise a sample of the overall population.

The use of Maximum Likelihood (ML) estimation method for conducting the CFA was primarily used due to restrictions concerning both samples size and access to statistical

packages that offered alternatives. The literature regarding the use of ML with non-normal samples indicates the possibility of bias and thus this must be considered when interpreting the current findings. In light of the possibility of bias, more emphasis was placed on alternative model fit statistics (e.g., RMSEA and CFI) during interpretation.

Initially, it was hoped to have a larger sample size to work with. A sizeable proportion of the data represented individuals with more than one admission. To satisfy the independence assumption only one admission could be used for the main analysis which shrunk the sample by 274. A larger sample may have enabled the results from the girls only CFA to have been conducted successfully and also allowed more specific exploration of results between group differences (e.g., Pacifika females vs NZ European females).

Conclusion

In conclusion, the current study raises serious questions as to the usefulness of the MAYSI-2 as a valid and reliable measure in the NZ context. While certain dimensions of the MAYSI-2 were sufficient by empirical standards, the areas in which the MAYSI-2 did not demonstrate adequate psychometric properties suggests the tool may not be useful for youth in secure care facilities within NZ. In light of the author's recommendations that the items in the MAYSI-2 not be modified or adapted (Grisso & Barnum, 2006), and taking into consideration the findings regarding its psychometric properties, the sensible option appears to be discontinuing the use of the MAYSI-2.

In relation to the second research question, the research established that NZ youth in secure care facilities present with varying degrees of mental and behavioural problems, as evidenced by the SACS, K6, and Suicide Risk Practice Tool. Specifically, substance use issues appeared to be far in excess of age equivalent community samples. Levels of 'psychological distress' were comparatively lower than both community samples of youth and adult samples,

raising questions as to the experience of ‘psychological distress’ by youth offenders, and also the usefulness of the K6 in detecting this. The number of youth deemed at risk of suicide was in line with international studies, though a lack of community sample comparisons of current suicidal ideation limited the ability to make firm conclusions.

A number of future research avenues are presented, including the development of a tool which is able to be used across the youth justice sector, not just in secure care facilities. Given the high rates of Māori youth involved with youth justice, the tool should be developed with appropriate cultural input in order to make it valid and relevant to the Māori population. It is hoped that this study draws attention to the need to validate psychometrics prior to implementation and widespread use, especially given both mental health service users and forensic populations have unique ethnic make-up, distinct from other countries.

Appendices

Appendix A: Mental health problems of youth in secure facilities: international literature

Table 1
Mental health problems of youth in secure facilities: international literature

Author/Date/ Location	Sample characteristics	Ethnicity	Measurement Tool(s) Used[1]	Percentage of sample meeting criteria for at least on mental health diagnosis	Major Findings of Interest from Study
(Allerton & Champion, 2003) New South Wales, Australia	N=242 223 males 19 females Age: 14-22	Aboriginal (42%), not reported (68%)	K-SADS-PL APSD CTQ	88%	Affective: MDD: 14%; mania: 15%; dysthymic disorder 13% SUD/AUD: Substance use disorder: 61% Anxiety: Separation AD: 26%; PTSD: 20%; OCD: 17%; GAD 11% Conduct: CD: 47%; aggression: 61%; ODD: 24% Psychosis: schizophrenia: 21% Suicide: attempted: 8.4% Trauma: any: 68% Other: ADHD:30%
(Indig et al., 2011) New South Wales, Australia	N=293 254 males 39 females Age: 13-21	Aboriginal (48%), not reported (62%)	SF-12 APS CTQ K-10	87%	Affective: any: 23.5%; MDD: 16.7%; bipolar: 4.4% SUD/AUD: any disorder: 63.5%; SUD: 49.3%; alcohol abuse disorder: 43.5% Anxiety: any: 31.7%; PTSD: 20.1; GAD: 7.2 Conduct: CD: 59.0% Psychosis: any: 5.5%; schizophrenia: 2.7% Suicide: attempted: 9.5%

					Trauma: any: 59.9% Other: ADHD: 29.7
(Pliszka et al., 2000) Texas, USA	N=50 45 males 5 females Age: 11-17	Not reported	DISC	76%	Affective: any: 42%; MDD: 20%; mania: 22% SUD/AUD: alcohol dependence: 28%; marijuana dependence: 46% Conduct: CD: 60%
(Hamerlynck et al., 2008) The Netherlands	N=216 0 males 216 females Age: 12-18	Dutch (57.1%), Surinamese (14.3%) Moroccan (8.1%), Dutch Antillean (4.3%), Turkish (1.4%), other (14.8%)	K-SADS-PL BDI DIS-Q CPTSD-RI	Not reported OR > 56%	Affective: depression: 33.3% SUD/AUD: Drug abuse: 51.6% Anxiety: PTSD: 21.0% Conduct: CD: 56.0%; ODD: 38.9% Suicide: Suicidality: 46.8% Other: ADHD: 20.8%
(Sawyer et al., 2010) Australia	N=159 132 males 27 females Age: 13-17	Non-Indigenous (64.2%), Indigenous (35.8%)	YSR SDQ YRBSS	50.3%	Anxiety: depressed/anxious combined scale: 19.5% Conduct: CD: 59.6% Suicide: attempted: 19.1%; ideation: 21.1; plan: 16.4%
(Goldstein et al., 2003) USA	N=232 0 males 232 females Age:12-18	Caucasians (58.2%), Latino (18.5%), African American (15.1%), Asian American (2.5%), other (5.6%)	MACI MAYSI-2 YSR	Not reported OR > 71.7%	Affective: depression: 63% SUD/AUD: any: 71.7% Anxiety: 56.3% Suicide: ideation: 36.2% Trauma: any: 52%

(Timmons-Mitchell et al., 1997) Ohio, USA	N=50 25 males 25 females Age: M: 15.9	Not reported	DISC SCL-90 MACI	Male = 100% Female = 100% Average diagnosis = 5	Affective: mood disorder: M= 72%, F= 88% SUD/AUD: M= 88%, F= 56% Anxiety: M= 52%, F= 72% Conduct: CD: M= 100%, F= 96% Psychosis: M= 16%, F= 12% Other: ADHD: M=76%, F=68%
(Shelton, 2001) Maryland, USA	N=312 252 males 60 females Age: 12-20	African American (57%), Caucasian (26%), other (17%)	DISC C-GAS	53%	Affective: any: 16.7% SUD/AUD: any: 37.2% Anxiety: any: 57.6% Conduct: disruptive behaviour disorders: 39.8% Psychosis: 32%
(Teplin et al., 2002) Illinois, USA	N=1829 1172 males 657 females Age: 10–18	African American (54%), non-Hispanic White (16.2%), Hispanic (28.7%), other (0.2%)	DISC-2.3	M= 66.3% F= 73.8%	Affective: any: M= 18.7%, F= 27.6%; MDD: M= 13.0%, F= 21.6% SUD/AUD: any: M= 50.7%, F= 46.8% Anxiety: any: M= 21.3%, F= 30.8% Conduct: CD: M= 37.8%, F= 40.6%; any disruptive behaviour disorder: M=41.4%, F= 45.6% Other: ADHD: M= 16.6%, F= 21.4%
(Bickel & Campbell, 2002) Australia	N=50 43 males 7 females Age: 12-18	Not reported	APS	98%	Affective: any: 46%; MDD= 30%; dysthymia: 16% Anxiety: any excluding PTSD: 32%; PTSD: 36% Conduct: CD: 98% Other: ADHD: 46%

(Domalanta et al., 2003) USA	N=1024 750 males 274 females Mean age: 14.9 for females, 15.4 for males	Males/females: white (18.5%/23.4%), black (42%/32.8%), Hispanic (35.3%/36.5%), other (4.2%/7.3%)	PHQ	60%	Affective: any: M= 26%, F= 31%; MDD: M= 10%, F= 10% SUD/AUD: SUD: M= 27%, F= 27%; AUD: M= 43%, F= 36% Anxiety: Panic: M= 5%, F= 8%; other: M= 8%, F= 12%
(Richards, 1996) Australia	N=100 80 males 20 females Age: 12-20 Mean age: 16.1 for females, 16.4 for males	Not reported	Semi structured diagnostic interview	100%	Affective: any: 25% Anxiety: PTSD: 4% Conduct: CD: 71% Psychosis: schizophrenia: 4% Other: ADHD: 4%
(Gretton & Clift, 2011) British Columbia, Canada	N=205 140 males 65 females Age: 12-20	European (37.9%), Aboriginal (31.4%), part Aboriginal (14.3%), other (13.6)	MAYSI-2 DISC-IV	M= 91.9% F= 100%	Affective: any: M= 5.8%, F= 7.4% SUD/AUD: any: M= 85.5%, F= 100% Anxiety: Any minus PTSD: M= 17.5%, F= 29.6%; PTSD: M= 1.7%, F= 13.0% Conduct: CD: M= 72.9%, F= 84.3% Trauma: physical abuse: M= 60.8%, F= 54.3%; sexual abuse: M= 21.2%, F= 42.4%; neglect: M= 51.0%, F= 62.9% Other: ADHD: M= 12.5%, F= 22.0%
(Atkins et al., 1999) South Carolina, USA	N= 75 71 males 4 females	African American (77.3%), White (22.7%)	DISC-PC-2.3	72%	Affective: any: 24% SUD/AUD: any: 20% Anxiety: any: 33.3%

	Age: 13-17 M=15.5				Conduct: any: 42.7%; CD: 40%; ODD: 15% Psychosis: 45.3% Other: ADHD: 1%
(Vreugdenhil et al., 2004) The Netherlands	N=204 204 males 0 females Age: 12-18	Surinamese (24%), Antillean (4%), Moroccan (22%), Turkish (7%), other (19%)	DISC-C DISC-IV DISC-2.3	Males= 90%	Affective: any: 6% SUD/AUD: any: 55% Anxiety: any: 9% Conduct: any: 75%; CD: 73%; ODD: 14% Psychosis: 34% Other: ADHD: 8%
(G. Wasserman et al., 2010) USA, multiple states	N=4961 3919 males 1042 females Age: M=16.3 SD=1.1	African American (33%), Hispanic (21.3%), White (38.6%), American Indian (4.7%), other (2.4%)	V-DISC	63.7%	Affective: any: 8.8% SUD/AUD: any: 47% Anxiety: any: 20.3% Conduct: any: 35.7% Suicide: attempted: 16.3%
(G. Wasserman et al., 2002) USA, multiple states	N= 292 292 males 0 females Age: M=17 SD=1.39	African American (54%) Hispanic (16%), White (28%), American other (2%)	V-DISC-IV	68.5%	Affective: any: 9.6%; MDD: 7.5% SUD/AUD: any: 50% Anxiety: any: 19.5%; PTSD: 5%; GAD: 2% Conduct: any: 32.5%; CD: 32%; ODD: 3.2% Suicide: attempted: 12.3% Other: ADHD: 2%
(Lader et al., 2003) England and Wales	N=590 Genders not reported	White (75%+)	SCID-II SCAN	At least 95%	SUD/AUD: M= 62-70%, F= 51% Psychosis: 8-10%

	Age: 16-20		CIS-R AUDIT		Suicide: attempted: 20% Other: personality disorders: 84-88%
(Harzke et al., 2012) Texas, USA	N=11,603 10,466 males 1,137 females Age: 71.3% ≥16	African American (30%), Hispanic (40%), Non-Hispanic white (19%), Other (11%)	Guided interview based on DSM-IV	98.3%	Affective: bipolar: 19.4%; depressive disorder: 12.6% SUD/AUD: any: 75.6% Anxiety: any: 6.4% Conduct: CD: 83.2%; ODD: 1.4% Psychosis: any: 2.0% Other: ADHD: 18.3%
(Ståhlberg et al., 2010) Sweden	N=100 92 males 8 females Age: 12-19 M=16 SD=1.5	Not reported	Clinical interview, BYI, YSR, ATAC, WAIS, WISC	73%	Affective: MDD: 20% SUD/AUD: SUD: 55% Anxiety: 18% Conduct: CD: 77% Psychosis: 3% Other: ADHD: 47%; mental retardation: 11%
(Chitsabesan et al., 2006) England and Wales	N=151 118 males 33 females Age: 13-18 M=15.7 SD=1.3	White (74%), black and ethnic minority (26%)	SNASA	Not reported	Affective: 19% SUD/AUD: 11%/6% Anxiety: 11%; PTSD: 11% Psychosis: 5% Suicide: self harm: 10%
(Archer et al., 2010) USA	N=1192 1082 males 110 females	African America (66.8%), White (27.3%), Hispanic	DSM Interview	Not reported	Affective: depression: 37.4%; M=36%, F=51.8% SUD/AUD: any: 63.1%; M=64.1%, F=50.9% Anxiety: any: 8.2%; M=6.9%, F=20.9%

	Age: M=16.0 SD=1.1	(4.3%), Asian (0.8%), other (0.8%)			Conduct: CD: 71.9%; M=74.6%, F=45.5% Psychosis: 0.9%; M=0.8%, F=1.8%
(Robertson et al., 2004) Mississippi, USA	N=482 292 males 190 females Age: M=15.3 SD=1.35	African American (65.4%), Caucasian (31.7%), other (2.9%)	APS, face-to-face sem structured interview	71-85%	Affective: MDD: M=12.6%, F=31.0% SUD/AUD: any: M=40.4%, F=27.7% Anxiety: GAD: M=11.1%, F=30.6%; PTSD: M=18.0%, F=41.0% Conduct: CD: M=50.5%, F=39.4% Psychosis: M=30.6%, F=29.5% Other: ADHD: M= 11.7%, F=28.0%
(Ghanizadeh et al., 2012) Iran	N=100 100 males Age: 12-19		K-SADS	~70%	Affective: MDD: 11% Anxiety: 0% Conduct: CD: 55%; ODD: 45% Psychosis: 2% Suicide: attempted: 12% Other: ADHD: 33%
(Dixon et al., 2004) Sydney, Australia	N=100 100 females Age: 13.5-19	Aboriginal (48%), White (33%), other (19%)	K-SADS-PL	99%	Affective: depression: 33% SUD/AUD: 85% Anxiety: PTSD: 20%; GAD: 5%; panic disorder: 9% Conduct: CD: 91% Trauma: physical abuse: 49%; sexual abuse: 50% Suicide: attempted: 46% Other: ADHD: 6%

[1] Refer to Appendix B for abbreviations

Appendix B: Abbreviations of mental health tests and disorders

Adolescent Psychopathology Scale (APS)
Alcohol Use Disorders Identification Test (AUDIT)
Antisocial Process Screening Device (APSD)
Anxiety Disorder (AD)
Attention deficit hyperactivity disorder (ADHD)
Autism-Tics, AD/HD and other Comorbidities inventory (ATAC)
Beck Depression Inventory (BDI)
Beck Youth Inventory (BYI)
Child Behaviour Checklist (CBCL)
Child Post-Traumatic Stress Disorder Reaction Index (CPTSD-RI)
Childhood Trauma Questionnaire (CTQ)
Children's Global Assessment Scale (CGAS)
Clinical Interview Schedule (CIS-R)
Conduct disorder (CD)
Diagnostic Interview Schedule for Children (DISC)
Dissociation Questionnaire (DIS-Q)
Generalized Anxiety Disorder (GAD)
Kessler Psychological Distress Scale (K10)
Kessler Psychological Distress Scale (K6)
Kiddie-Sads-Present and Lifetime Version (K-SADS-PL)
Major depressive disorder (MDD)
Millon Adolescent Clinical Inventory (MACI)
Minnesota Multiphasic Personality Inventory – Adolescent (MMPI-A)
Obsessive-Compulsive Disorder (OCD)
Oppositional defiant disorder (ODD)
Patient Health Questionnaire (PHQ)
Post-Traumatic Stress Disorder (PTSD)
Salford Needs Assessment Schedule for Adolescents (SNASA)
Schedules for the Clinical Assessment of Psychopathology (SCAN)
Short Form-12 Health Survey (SF-12)
Strengths and Difficulties Questionnaire (SDQ)
Structured Clinical Interview for DSM-IV (SCID-II)
Substance Use Disorders/Alcohol Use Disorder (SUD/AUD)
Substances and Choices Scale (SACS)
Symptom Checklist-90 (SCL-90)
Wechsler Adult Intelligence Test (WAIS)
Wechsler Intelligence Scale for Children (WISC)
Youth Risk Behavior Surveillance System (YRBSS)
Youth Self-Report (YSR)

Appendix C: The Massachusetts Youth Screening Instrument-Second Version (MAYSI-2; Grisso & Barnum, 2006)

For copyright reasons the Massachusetts Youth Screening Instrument-Second Version (MAYSI-2) cannot be reproduced here. However, a summary of the screening tool is provided below.

The MAYSI-2 is a brief self-report screening measure that can be administered to young people at any entry or transitional point within the youth justice system. The MAYSI-2 aims to identify youths aged 12 to 17 years old who are at risk for serious mental, emotional, and behavioural disorders and in need of clinical intervention within the juvenile justice setting (Grisso & Barnum, 2006). The MAYSI-2 is a 52-item screening tool which is administered within 24 to 48 hours after admission to a secure facility and consists of 7 scales of mental health or behavioural problems. It was designed to be administered, scored, and interpreted with ease, often by non-clinical personnel. It also has a short time frame, taking approximately 10-15 minutes to complete and 2-3 minutes to score. Respondents are asked to circle “yes” or “no” concerning whether the item has been true for them “within the past few months” on six of the scales, and “ever in your whole life” for one scale. The primary scales are as follows; Alcohol/Drug Use, Angry-Irritable, Depressed-Anxious, Somatic Complaints, Suicide Ideation, and Though Disturbance (for boys only). The seventh scale, Traumatic Experiences, provides information about the presence of trauma throughout their life. Example items for each scale are provided below.

Alcohol/Drug Use

Item 10: Have you done anything you wish you hadn’t when you were drunk or high?

Angry-Irritable

Item 2: Have you lost your temper easily, or had a “short fuse”?

Depressed-Anxious

Item 3: Have nervous or worried feelings kept you from doing things you want to do?

Somatic Complaints

Item 27: Have you felt shaky?

Suicidal Ideation

Item 26: Have you felt like killing yourself?

Thought Disturbance

Item 32: Have you been able to make other people do things just by thinking about it?

Traumatic experiences

Item 52: Have you ever seen someone severely injured or killed (in-person)

Appendix D: Substances and Choices Scale (SACS; Christie et al., 2007)

SUBSTANCES AND CHOICES SCALE

Name.....

Date of birth..... No.....

The SACS is only to be used by health professionals working with young people who are engaged in a treatment agency.

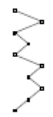
The questions in part A) and B) are about your use of alcohol and drugs over the last month. This does not include tobacco or prescribed medicine. Please answer every question as best you can, even if you are not certain. Tick only one box on each row.

A) How often did you use each of the following *in the last month?*	*Didn't use*	*Once a week or less*	*More than once a week*	*Most days or more*
1. Alcoholic drinks (e.g. beer, wine, spirits, premixes)	☐	☐	☐	☐
2. Cannabis (e.g. weed, marijuana)	☐	☐	☐	☐
3. Ecstasy and other party pills (e.g. 'E', Methadrone, BZP)	☐	☐	☐	☐
4. Hallucinogens (e.g. LSD, acid, mushrooms, ketamine)	☐	☐	☐	☐
5. Inhalants (e.g. glue, petrol, solvents, paint, nitrous)	☐	☐	☐	☐
6. Amphetamines (e.g. speed, 'P', ice, whiz)	☐	☐	☐	☐
7. Sedatives (e.g. sleeping pills, benzos, downers, valium)	☐	☐	☐	☐
8. Synthetic cannabinoids (smokable 'herbal highs')	☐	☐	☐	☐
9. Opiates (e.g. heroin, morphine, methadone, codeine)	☐	☐	☐	☐
10. Cocaine (e.g. coke, crack, blow)	☐	☐	☐	☐
11. Other drug. *Write name here*	☐	☐	☐	☐
12. Other drug. *Write name here*	☐	☐	☐	☐

B) Mark one box (on each row), on the basis of how things have been for you *over the last month.*	*Not True*	*Somewhat True*	*Certainly True*
1. I took alcohol or drugs when I was alone.	☐	☐	☐
2. I've thought I might be hooked or addicted to alcohol or drugs.	☐	☐	☐
3. Most of my free time has been spent getting hold of, taking, or recovering from alcohol or drugs.	☐	☐	☐
4. I've wanted to cut down on the amount of alcohol and drugs that I am using.	☐	☐	☐
5. My alcohol and drug use has stopped me getting important things done.	☐	☐	☐
6. My alcohol or drug use has led to arguments with the people I live with (family, flatmates or caregivers etc.).	☐	☐	☐
7. I've had unsafe sex or an unwanted sexual experience when taking alcohol or drugs.	☐	☐	☐
8. My performance or attendance at school (or at work) has been affected by my alcohol or drug use.	☐	☐	☐
9. I did things that could have got me into serious trouble (stealing, vandalism, violence etc) when using alcohol or drugs.	☐	☐	☐
10. I've driven a car while under the influence of alcohol or drugs (or have been driven by someone under the influence).	☐	☐	☐
	SACS difficulties scale		☐

YOUR SACS DIFFICULTIES MOUNTAIN RANGE

Connect the boxes with a straight line and turn the page up this way to see your SACS Difficulties Mountain Range like here.
Is your progress smooth or rocky?



C) Finally, how often have you used tobacco (e.g. cigarettes, cigars) *in the last month?*	*Never*	*Once a week or less*	*More than once a week*	*Most days or more*
	☐	☐	☐	☐

Date completed..... Clinician .....

Notes.....

.....

.....

SACSclinical2011 ©substancesandchoicesscale2011

Appendix E: Kessler-6 (K6) Psychological Distress Scale

This screen gives an indication of a young person's current distress and possible mental health issues.

Items of the Kessler-6 (K6) Non-Specific Psychological Distress Scales.

During the last two weeks about how often did you feel…	Rating (0-4)	Notes
So sad nothing could cheer you up?		
Worried or frightened?		
Restless or stressed?		
Hopeless?		
That everything was an effort?		
Worthless?		

Rating Scale (e.g., most of the time)
0 – none
1 – little
2 – some
3 – most
4 – all

Appendix F: Suicide Risk Practice Tool

For recognition of any risk of suicide

How do you see the future?

Do you ever feel life is not worth living?

Have you ever thought you would like to end it all?

Is the young person thinking about suicide or significant self harm?

YES or NO (circle one)

Appendix G: Additional Output from MAYSI-2 CFA

Table 1

Standardised (pattern coefficients) and unstandardized factor loadings for five MAYSI-2 scales (n=527)

MAYSI-2 Item		MAYSI-2 Scale	Estimate (factor loading)	Standardised loading	Bootstrap p value
Item42	<--	AI	.590	.477	.001
Item39	<--	AI	.844	.646	.001
Item35	<--	AI	1.000	.802	...
Item13	<--	AI	.780	.649	.001
Item8	<--	AI	.452	**.387**	.001
Item7	<--	AI	.553	.430	.001
Item6	<--	AI	.658	.508	.001
Item2	<--	AI	.788	.607	.001
Item44	<--	AI	.593	.460	.001
Item45	<--	ADU	.896	.601	.001
Item40	<--	ADU	.894	.645	.001
Item37	<--	ADU	.840	.571	.001
Item33	<--	ADU	.795	.541	.001
Item24	<--	ADU	.772	.534	.001
Item23	<--	ADU	1.000	.717	...
Item19	<--	ADU	.663	.475	.001
Item10	<--	ADU	.848	.564	.002
Item47	**<--**	**DA**	**.230**	**.173**	**.015**
Item41	<--	DA	.804	.411	.001
Item34	<--	DA	.794	.435	.001

MAYSI-2 Item		MAYSI-2 Scale	Estimate (factor loading)	Standardised loading	Bootstrap p value
Item21	<--	DA	.706	.453	.001
Item17	<--	DA	.875	.512	.001
Item14	<--	DA	.640	.454	.001
Item3	<--	DA	1.000	.490	...
Item51	<--	DA	.903	.501	.001
Item11	<--	SI	.950	.759	.001
Item16	<--	SI	1.000	.785	...
Item18	<--	SI	.929	.816	.001
Item22	<--	SI	.994	.874	.001
Item27	<--	SC	.765	.540	.001
Item28	<--	SC	1.000	.653	...
Item29	<--	SC	.824	.623	.002
Item30	<--	SC	.655	.464	.001
Item31	<--	SC	.793	.575	.001
Item43	<--	SC	.610	.443	.001
Item35	**<--**	**DA**	**-.353**	**-.166**	**.035**
Item47	<--	SI	.506	.503	.001

Table 2

Co-variance matrix for five MAYSI-2 scales/latent variables (n=527)

	ADU	AI	DA	SI	SC
ADU	0.111	-	-	-	-
AI	0.061	0.147	-	-	-
DA	0.021	0.057	0.05	-	-
SI	0.018	0.059	0.038	0.088	-
SC	0.045	0.089	0.057	0.045	0.105

All significant at $p < .0001$ level

Table 3

Latent variable correlations for five MAYSI-2 scales/latent variables (n=527)

	SC	SI	DA	ADU	AI
SC	1				
SI	0.466	1			
DA	0.783	0.577	1		
ADU	0.421	0.18	0.286	1	
AI	0.715	0.523	0.666	0.476	1

Table 4

Correlation coefficients (structure coefficients) for five MAYSI-2 scales (n=527)

	SC	SI	DA	ADU	AI
Item43	**0.443**	0.206	0.347	0.186	0.317
Item31	**0.575**	0.268	0.45	0.242	0.411
Item30	**0.464**	0.216	0.363	0.195	0.332
Item29	**0.623**	0.29	0.487	0.262	0.445
Item28	**0.653**	0.304	0.511	0.275	0.467
Item27	**0.54**	0.252	0.423	0.227	0.386
Item22	0.407	**0.874**	0.504	0.157	0.457
Item18	0.38	**0.816**	0.471	0.147	0.427
Item16	0.365	**0.785**	0.452	0.141	0.41
Item11	0.353	**0.759**	0.438	0.136	0.397
Item51	0.392	0.289	**0.501**	0.144	0.333
Item3	0.384	0.283	**0.49**	0.141	0.327
Item14	0.355	0.262	**0.454**	0.13	0.302
Item17	0.401	0.295	**0.512**	0.147	0.341
Item21	0.354	0.261	**0.453**	0.13	0.301
Item34	0.34	0.251	**0.435**	0.125	0.289
Item41	0.321	0.237	**0.411**	0.118	0.273
Item47	0.369	**0.603**	**0.463**	0.14	0.378
Item10	0.238	0.101	0.162	**0.564**	0.269
Item19	0.2	0.085	0.136	**0.475**	0.226
Item23	0.302	0.129	0.206	**0.717**	0.342
Item24	0.225	0.096	0.153	**0.534**	0.254
Item33	0.228	0.097	0.155	**0.541**	0.258
Item37	0.24	0.102	0.163	**0.571**	0.272
Item40	0.271	0.116	0.185	**0.645**	0.307
Item45	0.253	0.108	0.172	**0.601**	0.286
Item44	0.329	0.24	0.306	0.219	**0.46**
Item2	0.434	0.317	0.404	0.289	**0.607**
Item6	0.363	0.266	0.338	0.242	**0.508**
Item7	0.308	0.225	0.286	0.205	**0.43**
Item8	0.277	0.202	0.258	0.184	**0.387**
Item13	0.464	0.339	0.432	0.309	**0.649**
Item35	0.444	0.324	**0.368**	0.334	**0.692**

Item39	0.462	0.338	0.43	0.308	**0.646**
Item42	0.341	0.249	0.317	0.227	**0.477**
	N	Y	N	Y	**N**

Note: Bold numbers show loading of the measurement items on the constructs to which they are assigned to in the CFA.

Table 5

Standardised (pattern coefficients) and unstandardized factor loadings for seven MAYSI-2 scales (boys only; n=432)

MAYSI-2 Item		MAYSI-2 Scale	Estimate (factor loading)	Standardized factor loading	Bootstrap p value
Item42	<---	AI	.604	.473	.001
Item39	<---	AI	.890	.652	.001
Item35	<---	AI	1.000	.785	...
Item13	<---	AI	.770	.634	.001
Item8	<---	AI	.445	.372	.001
Item7	<---	AI	.506	.377	.001
Item6	<---	AI	.690	.514	.001
Item2	<---	AI	.820	.604	.001
Item44	<---	AI	.585	.431	.001
Item45	<---	ADU	.958	.616	.001
Item40	<---	ADU	.926	.640	.001
Item37	<---	ADU	.831	.539	.001
Item33	<---	ADU	.808	.525	.001
Item24	<---	ADU	.788	.526	.001
Item23	<---	ADU	1.000	.691	...
Item19	<---	ADU	.683	.479	.001
Item10	<---	ADU	.877	.559	.001
Item47	<---	DA	**.195**	**.140**	**.142**
Item41	<---	DA	.800	.381	.001
Item34	<---	DA	.875	.452	.001
Item21	<---	DA	.617	.377	.002
Item17	<---	DA	.884	.473	.001
Item14	<---	DA	.473	.336	.001
Item3	<---	DA	1.000	.451	...
Item51	<---	DA	.534	.277	.003
Item11	<---	SI	.909	.742	.001
Item16	<---	SI	1.000	.778	...
Item18	<---	SI	.876	.783	.002
Item22	<---	SI	.965	.862	.002
Item27	<---	SC	.759	.542	.001
Item28	<---	SC	1.000	.658	...
Item29	<---	SC	.759	.606	.001
Item30	<---	SC	.621	.450	.001

MAYSI-2 Item		MAYSI-2 Scale	Estimate (factor loading)	Standardized factor loading	Bootstrap p value
Item31	<---	SC	.700	.526	.002
Item43	<---	SC	.598	.445	.002
Item35	<---	DA	**-.352**	**-.154**	**.142**
Item47	<---	SI	.505	.496	.001
Item9	<---	TD	.647	.592	.001
Item20	<---	TD	.701	.622	.001
Item25	<---	TD	.255	.291	.001
Item26	<---	TD	1.000	.679	...
Item32	<---	TD	.262	.311	.001
Item46	<---	TE	.564	.374	.001
Item48	<---	TE	.969	.628	.001
Item49	<---	TE	1.000	.648	...
Item52	<---	TE	.902	.582	.001
Item51	<---	TE	.384	.316	.005

Table 6

Co-variance matrix for seven MAYSI-2 scales/latent variables (boys only; n=432)

	ADU	AI	DA	SI	SC	TD	TE
ADU	0.102	-	-	-	-	-	-
AI	0.055	0.134	-	-	-	-	-
DA	0.015b	0.051	0.042	-	-	-	-
SI	0.015a	0.053	0.03	0.078	-	-	-
SC	0.043	0.081	0.055	0.042	0.106	-	-
TD	0.03	0.54	0.034	0.032	0.053	0.074	-
TE	0.051	0.063	0.037	0.027	0.059	0.039	0.104

All significant at $p < .001$ level with exception of a and b ($p = .004$ and .002 respectively)

Table 7

Latent variable correlations for seven MAYSI-2 scales/latent variables (boys only; n=432)

	TE	TD	SC	SI	DA	ADU	AI
TE	1.000						
TD	.449	1.000					
SC	.566	.606	1.000				
SI	.303	.419	.466	1.000			
DA	.565	.610	.831	.536	1.000		
ADU	.500	.341	.420	.168	.235	1.000	
AI	.534	.544	.682	.516	.680	.469	1.000

Table 8

Correlation coefficients (structure coefficients) for seven MAYSI-2 scales (boys only; n=432)

	TE	TD	SC	SI	DA	ADU	AI
Item52	**.582**	.262	.330	.177	.329	.291	.311
Item49	**.648**	.291	.367	.196	.367	.324	.346
Item48	**.628**	.282	.356	.190	.355	.314	.336
Item46	**.374**	.168	.212	.113	.212	.187	.200
Item32	.140	**.311**	.188	.130	.189	.106	.169
Item26	.305	**.679**	.411	.285	.414	.232	.369
Item25	.131	**.291**	.176	.122	.177	.099	.158
Item20	.279	**.622**	.376	.261	.379	.212	.338
Item9	.266	**.592**	.359	.248	.361	.202	.322
Item43	.252	.269	**.445**	.207	.370	.187	.304
Item31	.298	.318	**.526**	.245	.437	.221	.359
Item30	.255	.273	**.450**	.210	.374	.189	.307
Item29	.343	.367	**.606**	.283	.504	.255	.414
Item28	.372	.398	**.658**	.306	.546	.276	.449
Item27	.307	.328	**.542**	.253	.450	.228	.370
Item22	.261	.361	.402	**.862**	.462	.145	.445
Item18	.237	.328	.365	**.783**	.419	.131	.404
Item16	.236	.326	.363	**.778**	.417	.131	.401
Item11	.225	.311	.346	**.742**	.397	.125	.383
Item51	**.473**	.311	.409	.244	**.456**	.223	.358
Item3	.255	.275	.375	.242	**.451**	.106	.307
Item14	.190	.205	.279	.180	**.336**	.079	.229
Item17	.267	.288	.393	.253	**.473**	.111	.322
Item21	.213	.230	.313	.202	**.377**	.089	.256
Item34	.256	.276	.376	.242	**.452**	.106	.308
Item41	.215	.232	.316	.204	**.381**	.090	.259
Item47	.230	.294	.348	**.571**	**.406**	.116	.351
Item10	.280	.191	.235	.094	.132	**.559**	.262
Item19	.240	.163	.201	.080	.113	**.479**	.225
Item23	.346	.236	.290	.116	.163	**.691**	.324
Item24	.263	.180	.221	.088	.124	**.526**	.247
Item33	.263	.179	.220	.088	.124	**.525**	.246
Item37	.270	.184	.226	.091	.127	**.539**	.253
Item40	.320	.219	.269	.108	.151	**.640**	.300
Item45	.308	.210	.259	.103	.145	**.616**	.289
Item44	.230	.235	.294	.222	.293	.202	**.431**
Item2	.323	.329	.412	.312	.411	.283	**.604**
Item6	.275	.280	.351	.265	.350	.241	**.514**
Item7	.201	.205	.257	.194	.256	.177	**.377**
Item8	.199	.203	.254	.192	.253	.175	**.372**
Item13	.339	.345	.433	.327	.431	.297	**.634**
Item35	.332	.333	.407	.322	**.380**	.332	**.680**

	TE	TD	SC	SI	DA	ADU	AI
Item39	.349	.355	.445	.336	.444	.306	**.652**
Item42	.253	.257	.323	.244	.322	.222	**.473**
	Y	N	N	Y	N	Y	N

Note: Bold numbers show loading of the measurement items on the constructs to which they are assigned to in the CFA.

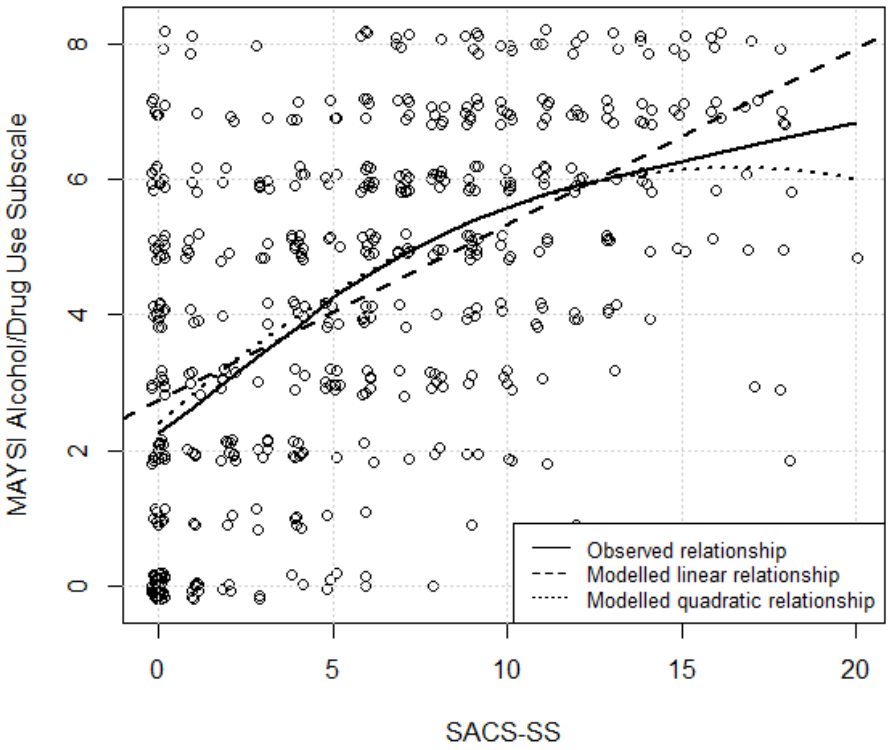


Figure 1. Scatter plot with modelled squared term (quadratic) and modelled linear relationship overlays.

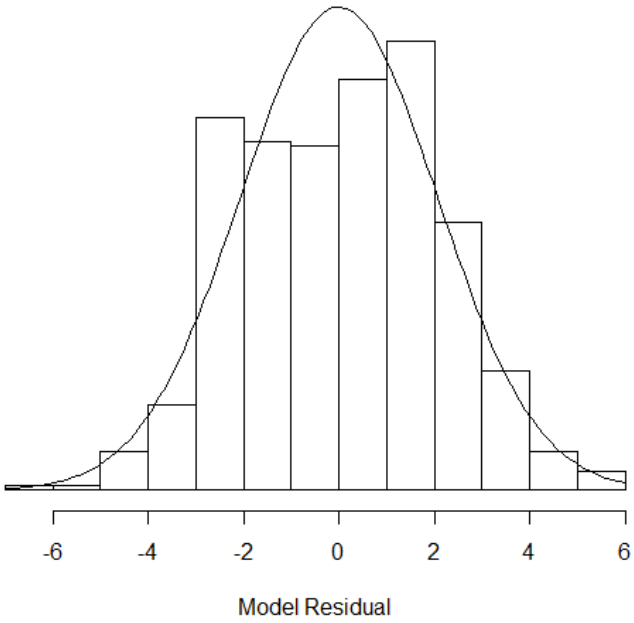


Figure 2. Normal distribution of residuals.

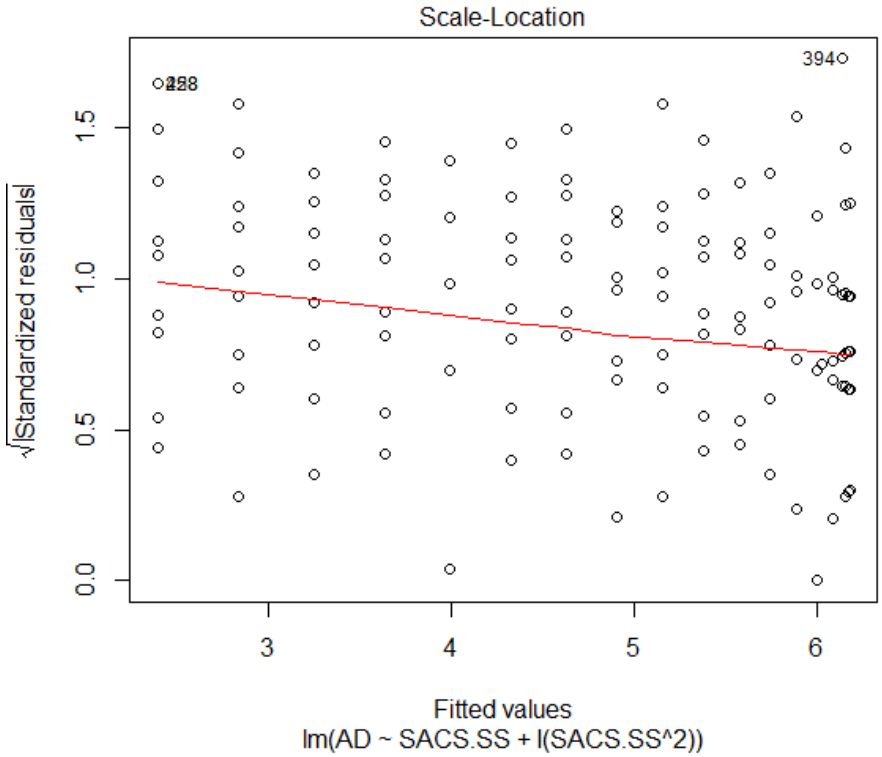


Figure 3. Scatter plot illustrating no issues with homoscedasticity.

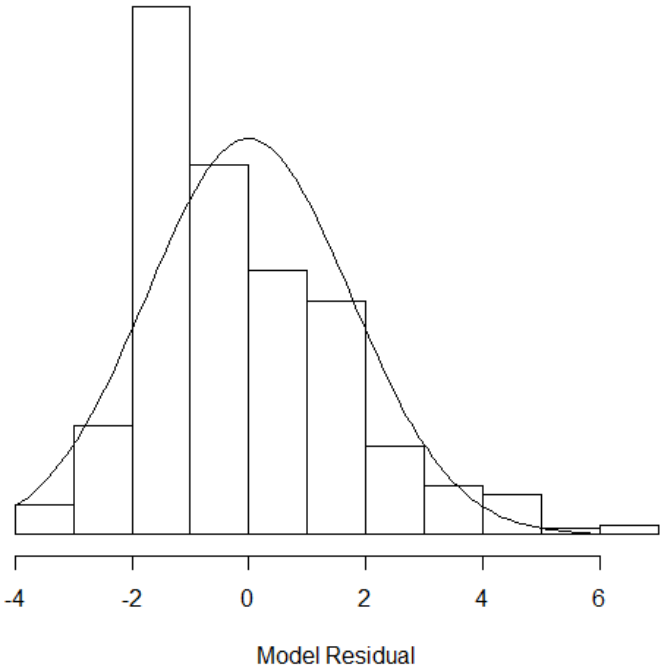


Figure 4. Normal distribution illustrating some concerns with assumption of normality.

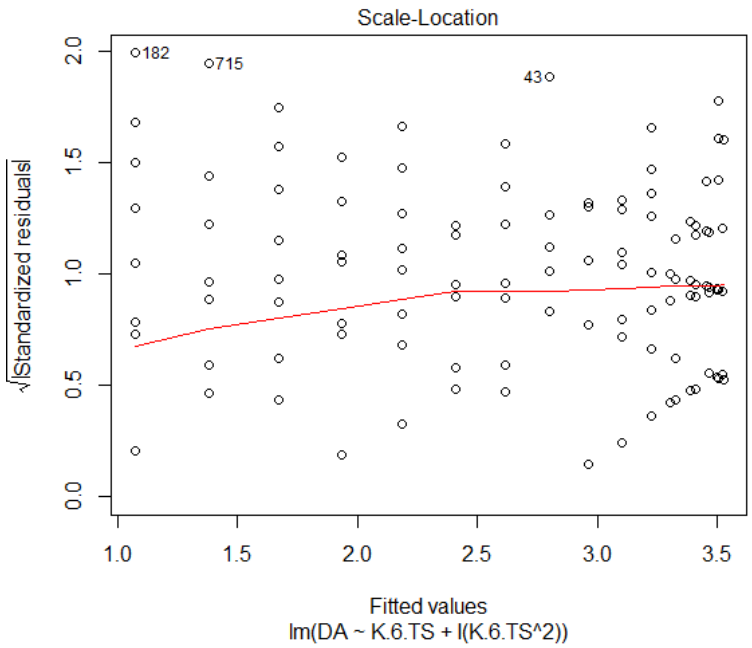


Figure 5. Scatter plot illustrating no issues with homoscedasticity.

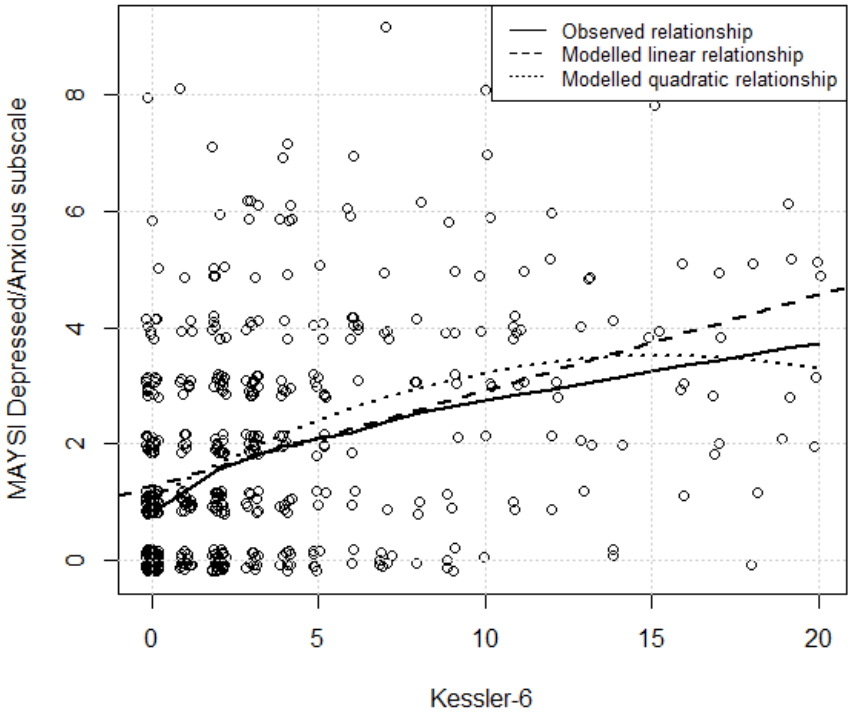


Figure 6. Scatter plot with modelled squared term (quadratic) and modelled linear relationship overlays.

Appendix H: International MAYSI-2 studies and results

Table 1
MAYSI-2 studies and results: means and standard deviations

STUDY	Gender	ADU		AI		DA		SC		SI		TD		TE	
		M	(SD)	M	(SD)	M	(SD)	M	(SD)	M	(SD)	M	(SD)	M	(SD)
(Colins et al., 2014) N=955 The Netherlands	Males 955	1.19	(1.96)	1.96	(2.15)	1.17	(1.50)	1.74	(1.47)	0.24	(0.80)	0.32	(0.69)	1.51	(1.39)
(Lennox et al., 2014) N=206 UK	Males 206	**5.15**	(2.55)	**5.06**	(2.81)	2.17	(1.97)	2.41	(1.78)	0.93	(1.56)	0.63	(0.91)	2.29	(1.53)
(Stathis et al., 2008) N=164 Australia	Male 124 Female 40	**4.0** **4.0**	(2.4) (2.2)	3.2 3.45	(2.5) (2.5)	1.4 2.65	(1.8) (2.4)	1.7 2.43	(1.8) (1.9)	0.5 1.15	(1.0) (1.6)	-	-	-	-
(Leenarts et al., 2016) N = 446 Switzerland	Male 297 Female 149	2.56 2.51	(2.75) (2.67)	4.28 4.84	(2.75) (2.55)	2.44 **4.01**	(2.12) (2.45)	1.54 2.80	(1.45) (1.76)	1.25 **2.22**	(1.66) (1.90)	-	-	-	-

Current study N=529 New Zealand	Male 434 Female 95	**4.23** **4.39**	(2.43) (2.61)	3.39 4.59	(2.58) (2.79)	1.70 2.53	(1.77) (2.39)	1.66 2.39	(1.69) (1.90)	0.58 1.20	(1.27) (1.83)	0.50 NA	(0.92) NA	1.95 2.06	(1.56) (1.61)
(Gilbert, Grande, Hallman, & Underwood, 2014) N=4,009 USA	Male 3,491 Female 518	3.51	(2.60)	2.67	(2.58)	1.31	(1.57)	1.85	(1.8)	0.30	(0.90)	0.31	(0.72)	1.77	(1.54)
(McCoy, Vaughn, Maynard, & Salas-Wright, 2014) N=836 USA	Male 660 Female 176	3.52	(2.96)	3.64	(2.89)	2.57	(2.21)	2.35	(1.92)	0.94	(1.31)	**1.03**	(1.24)	**3.01** **3.04**	(1.45) (1.67)
(Cauffman, Lexcen, Goldweber, Shulman, & Grisso, 2007) N=433 USA	Male 276 Female 157	3.11	-	4.56	-	2.68	-	**3.03**	-	0.84	-	-	-	-	-

(Archer et al.,	Male	1.5	(2.1)	3.0	(2.6)	1.5	(1.8)	1.8	(1.9)	0.7	(1.4)	0.05	(0.9)	1.5	(1.5)
2004)	548	2.1	(2.5)	4.5	(2.8)	2.5	(2.1)	2.9	(2.0)	1.2	(1.7)	NA	NA	2.0	(1.7)
N=704	Female														
USA	156														
(Coker et al.,	Male	0.72	(1.39)	2.37	(2.27)	1.21	(1.48)	1.54	(1.39)	0.11	(0.51)	0.25	(0.58)	1.31	(1.29)
2013)	3,734	0.55	(1.31)	3.1	(2.60)	1.71	(1.88)	1.94	(1.61)	0.35	(0.99)	NA	NA	1.17	(1.27)
N=5,205	Female														
USA	685														
(Grisso et al.,	Male	2.41	(2.49)	3.62	(2.69)	1.97	(2.01)	1.91	(1.86)	0.49	(1.09)	0.5	(0.9)	2.0	(1.5)
2001)	857	2.51	(2.62)	4.60	(2.64)	2.73	(2.29)	2.72	(2.03)	1.31	(1.66)	NA	NA	-	-
N=857	Female														
USA	408														
(Archer et al.,	Male	1.80	(2.29)	2.61	(2.65)	1.30	(1.78)	1.46	(1.75)	0.42	(1.09)	0.35	(0.73)	1.42	(1.50)
2010)	1,082	2.44	(2.76)	3.44	(2.81)	1.54	(1.89)	2.11	(2.02)	0.63	(1.31)	NA	NA	1.88	(1.71)
N=1,192	Female														
USA	110														
(Nordness et al.,	Male	2.3	(2.5)	3.2	(2.8)	1.6	(1.9)	2.1	(1.8)	0.6	(1.4)	0.4	(0.9)	-	-
2002)	150	2.5	(2.3)	4.1	(2.5)	2.8	(2.3)	**3.0**	(1.9)	1.3	(1.8)	NA	NA	-	-
N=204	Female														
USA	40														

Note: Means in bold represent average scores above the cut-off range.

References

Abram, K. M., Choe, J. Y., Washburn, J. J., Teplin, L. A., King, D. C., & Dulcan, M. K. (2008). Suicidal ideation and behaviors among youths in juvenile detention. *Journal of the American Academy of Child & Adolescent Psychiatry, 47*(3), 291-300.

Abram, K. M., Teplin, L. A., Charles, D. R., Longworth, S. L., McClelland, G. M., & Dulcan, M. K. (2004). Posttraumatic stress disorder and trauma in youth in juvenile detention. *Archives of General Psychiatry, 61*(4), 403-410.

Abram, K. M., Teplin, L. A., McClelland, G. M., & Dulcan, M. K. (2003). Comorbid psychiatric disorders in youth in juvenile detention. *Archives of General Psychiatry, 60*(11), 1097-1108.

Abram, K. M., Zwecker, N. A., Welty, L. J., Hershfield, J. A., Dulcan, M. K., & Teplin, L. A. (2015). Comorbidity and Continuity of Psychiatric Disorders in Youth After Detention: A Prospective Longitudinal Study. *JAMA psychiatry, 72*(1), 84-93.

Allerton, M., & Champion, U. (2003). *2003 NSW Young People in Custody Health Survey: Key Findings Report*: NSW Department of Juvenile Justice.

Anastasi, A. (1968). Psychological testing.

Anderson, J. C., Williams, S., McGee, R., & Silva, P. A. (1987). DSM-III disorders in preadolescent children: Prevalence in a large sample from the general population. *Archives of General Psychiatry, 44*(1), 69-76.

Anthenelli, R. M. (2010). Focus on: Comorbid mental health disorders. *Alcohol Research & Health, 33*(1-2), 109.

Arbuckle, J. L., & Wothke, W. (1999). *Amos 4.0 user's guide*: SmallWaters Corporation Chicago, IL.

Archer, R. P., Simonds-Bisbee, E. C., Spiegel, D. R., Handel, R. W., & Elkins, D. E. (2010). Validity of the Massachusetts Youth Screening Instrument–2 (MAYSI–2) Scales in Juvenile Justice Settings. *Journal of personality assessment, 92*(4), 337-348.

Archer, R. P., Stredny, R. V., Mason, J. A., & Arnau, R. C. (2004). An examination and replication of the psychometric properties of the Massachusetts Youth Screening Instrument-(MAYSI-2) among adolescents in detention settings. *Assessment, 11*(4), 290-302.

Association, A. e. R., Association, A. P., & Education, N. C. o. M. i. (1999). *Standards for educational and psychological testing*: American Educational Research Association.

Association, A. P. (2000). *Diagnostic and statistical manual of mental disorders, text revision (DSM-IV-TR)*: American Psychiatric Association.

Association, A. P. (2013). *Diagnostic and statistical manual of mental disorders, (DSM-5®)*: American Psychiatric Pub.

Atkins, D. L., Pumariega, A. J., Rogers, K., Montgomery, L., Nybro, C., Jeffers, G., & Sease, F. (1999). Mental health and incarcerated youth. I: Prevalence and nature of psychopathology. *Journal of Child and Family Studies, 8*(2), 193-204.

Barker-Collo, S. L. (2003). Culture and validity of the Symptom Checklist-90-Revised and Profile of Mood States in a New Zealand student sample. *Cultural Diversity and Ethnic Minority Psychology, 9*(2), 185.

Benson, J., & Fleishman, J. A. (1994). The robustness of maximum likelihood and distribution-free estimators to non-normality in confirmatory factor analysis. *Quality and Quantity, 28*(2), 117-136.

Best, C. (2013). Violent girls and the youth justice system: Why current systems are failing this at risk population. *Pub. Int. LJNZ*, 54.

Bickel, R., & Campbell, A. (2002). Mental health of adolescents in custody: the use of the 'Adolescent Psychopathology Scale'in a Tasmanian context. *Australian and New Zealand journal of psychiatry, 36*(5), 603-609.

Bird, H. R. (1996). Epidemiology of Childhood Disorders in a Cross–Cultural Context. *Journal of Child Psychology and Psychiatry, 37*(1), 35-49.

Bird, H. R., Davies, M., Fisher, P., Narrow, W. E., Jensen, P. S., Hoven, C., . . . Dulcan, M. K. (2000). How specific is specific impairment? *Journal of the American Academy of Child & Adolescent Psychiatry, 39*(9), 1182-1189.

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*(3), 19-23.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*: Wiley.

Bonta, J., & Andrews, D. A. (2007). Risk-need-responsivity model for offender assessment and rehabilitation. *Rehabilitation, 6*, 1-22.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37*(1), 62-83.

Browne, M. W., Cudeck, R., Bollen, K. A., & Long, J. S. (1993). Alternative ways of assessing model fit. *Sage focus editions, 154*, 136-136.

Butler, M. A., Loney, B. R., & Kistner, J. (2007). The Massachusetts Youth Screening Instrument as a predictor of institutional maladjustment in severe male juvenile offenders. *Criminal Justice and Behavior*.

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. In: American Psychological Association.

Caldwell, Z. (2009). The New Zealand Youth Offending Strategy-Te Haonga. In: Wellington: Ministry of Justice.

Capaldi, D. M. (1992). Co-occurrence of conduct problems and depressive symptoms in early adolescent boys: II. A 2-year follow-up at Grade 8. *Development and psychopathology, 4*(01), 125-144.

Capaldi, D. M., & Stoolmiller, M. (1999). Co-occurrence of conduct problems and depressive symptoms in early adolescent boys: III. Prediction to young-adult adjustment. *Development and psychopathology, 11*(01), 59-84.

Casswell, M., French, P., & Rogers, A. (2012). Distress, defiance or adaptation? A review paper of at-risk mental health states in young offenders. *Early intervention in psychiatry, 6*(3), 219-228.

Cauffman, E. (2004). A statewide screening of mental health symptoms among juvenile offenders in detention. *Journal of the American Academy of Child & Adolescent Psychiatry, 43*(4), 430-439.

Cauffman, E., Feldman, S., Watherman, J., & Steiner, H. (1998). Posttraumatic stress disorder among female juvenile offenders. *Journal of the American Academy of Child & Adolescent Psychiatry, 37*(11), 1209-1216.

Cauffman, E., Lexcen, F. J., Goldweber, A., Shulman, E. P., & Grisso, T. (2007). Gender differences in mental health symptoms among delinquent and community youth. *Youth Violence and Juvenile Justice, 5*(3), 287-307.

Cauffman, E., & MacIntosh, R. (2006). A Rasch Differential Item Functioning analysis of the Massachusetts Youth Screening Instrument identifying race and gender differential item functioning among juvenile offenders. *Educational and Psychological Measurement, 66*(3), 502-521.

Chan, S. M., & Fung, T. C. T. (2014). Reliability and validity of K10 and K6 in screening depressive symptoms in Hong Kong adolescents. *Vulnerable Children and Youth Studies, 9*(1), 75-85.

Chapman, J. F., & Ford, J. D. (2008). Relationships between suicide risk, traumatic experiences, and substance use among juvenile detainees. *Archives of Suicide Research, 12*(1), 50-61.

Chitsabesan, P., Kroll, L., Bailey, S., Kenning, C., Sneider, S., MacDONALD, W., & Theodosiou, L. (2006). Mental health needs of young offenders in custody and in the community. *The British Journal of Psychiatry, 188*(6), 534-540.

Christie, G. (2011). *Screening for substance use problems in New Zealand adolescents: Applications in comorbidity and treatment.* ResearchSpace@ Auckland,

Christie, G., Marsh, R., Sheridan, J., Wheeler, A., Suaalii-Sauni, T., Black, S., & Butler, R. (2007). The Substances and Choices Scale (SACS)-The development and testing of a new alcohol and other drug screening and outcome measurement instrument for young people. *Addiction, 102*(9), 1390-1398.

Christie, G., Marsh, R., Sheridan, J., Wheeler, A., Suaalii-Sauni, T., Black, S., & Butler, R. (2010). *THE SUBSTANCES AND CHOICES SCALE MANUAL*.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284.

Clark, T., Fleming, T., Bullen, P., Crengle, S., Denny, S., Dyson, B., . . . Sheridan, J. (2013). Health and well-being of secondary school students in New Zealand: Trends between 2001, 2007 and 2012. *Journal of paediatrics and child health, 49*(11), 925-934.

Clark, T., Fleming, T., Bullen, P., Denny, S., Crengle, S., Dyson, B., . . . Utter, J. (2013). *Youth'12 Overview: The health and wellbeing of New Zealand secondary school students in 2012*. Auckland, New Zealand.

Coker, K. L., Wernsman, J., Ikpe, U. N., Brooks, J. S., Bushell, L. L., & Kahn, B. A. (2013). Using the Massachusetts Youth Screening Instrument–Version 2 on a Community

Sample of African American and Latino/a Juvenile Offenders to Identify Mental Health and Substance Abuse Treatment Needs. *Criminal Justice and Behavior*, 0093854813505565.

Cole, D. A., & Carpentieri, S. (1990). Social status and the comorbidity of child depression and conduct disorder. *Journal of consulting and clinical psychology*, *58*(6), 748.

Colins, O., Grisso, T., Mulder, E., & Vermeiren, R. (2015). The relation of standardized mental health screening and categorical assessment in detained male adolescents. *European child & adolescent psychiatry*, *24*(3), 339-349.

Colins, O., Grisso, T., Vahl, P., Guy, L., Mulder, E., Hornby, N., . . . Vermeiren, R. (2014). Standardized Screening for Mental Health Needs of Detained Youths from Various Ethnic Origins: The Dutch Massachusetts Youth Screening Instrument-Second Version (MAYSI-2). *Journal of Psychopathology and Behavioral Assessment*, 1-12.

Colins, O., Vermeiren, R., Vreugdenhil, C., van den Brink, W., Doreleijers, T., & Broekaert, E. (2010). Psychiatric disorders in detained male adolescents: a systematic literature review. *Canadian journal of psychiatry. Revue canadienne de psychiatrie*, *55*(4), 255-263.

Commission, M. H. (2011). National Indicators 2011: Measuring mental health and addiction in New Zealand. *Wellington: Mental Health Commission*.

Commission, M. H. (2012). National Indicators 2012: Measuring mental health and addiction in New Zealand. *Wellington: Mental Health Commission*.

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis*: Psychology Press.

Crengle, S., Clark, T., Robinson, E., Bullen, P., Dyson, B., Denny, S., . . . Utter, J. (2012). The health and wellbeing of Māori New Zealand secondary school students in 2012. Te Ara Whakapiki Taitamariki: Youth'12. In: The Adolescent Health Research Group (2013) New Zealand: The University of Auckland.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika, 16*(3), 297-334.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin, 52*(4), 281.

Cruise, K. R., Marsee, M. A., Dandreaux, D. M., & DePrato, D. K. (2007). Mental health screening of female juvenile offenders: Replication of a subtyping strategy. *Journal of Child and Family Studies, 16*(5), 615-625.

Cuellar, A. E., Markowitz, S., & Libby, A. M. (2004). Mental health and substance abuse treatment and juvenile crime. *Journal of Mental Health Policy and Economics*, 59-68.

Dembo, R., Williams, L., & Schmeidler, J. (1993). Gender differences in mental health service needs among youths entering a juvenile detention center. *Journal of Prison & Jail Health*.

Demirbaş-Çakır, E., Şişmanlar, Ş. G., Çakır, U., Duymaz, M. K., Karakaya, I., Coşkun, A., & Ağaoğlu, B. (2015). Adolescent Delinquency and Psychopathology: Is Maysi-2 a Valid Instrument to Assess Psychopathology Risk among Turkish Adolescents? *Acta Medica Anatolia, 3*(4), 119-128.

Devlieger, E., & Verschuere, B. (2010). Youth delinquents and mental complaints: a study with the revised Massachusetts youth screening instrument (MAYSI-2) with Flemish juvenile boys. *PSYCHOLOGIE & GEZONDHEID, 38*(3), 138-142.

Dierkhising, C. B., Ko, S. J., Woods-Jaeger, B., Briggs, E. C., Lee, R., & Pynoos, R. S. (2013). Trauma histories among justice-involved youth: Findings from the National Child Traumatic Stress Network. *European journal of psychotraumatology, 4*.

Dixon, A., Howie, P., & Starling, J. (2004). Psychopathology in female juvenile offenders. *Journal of Child Psychology and Psychiatry, 45*(6), 1150-1158.

Dölitzsch, C., Leenarts, L. E., Schmeck, K., Fegert, J. M., Grisso, T., & Schmid, M. (2017). Diagnostic performance and optimal cut-off scores of the Massachusetts youth screening instrument-second version in a sample of Swiss youths in welfare and juvenile justice institutions. *BMC psychiatry, 17*(1), 61.

Domalanta, D. D., Risser, W. L., Roberts, R. E., & Risser, J. M. H. (2003). Prevalence of depression and other psychiatric disorders among incarcerated youths. *Journal of the American Academy of Child & Adolescent Psychiatry, 42*(4), 477-484.

Drapeau, A., Beaulieu-Prévost, D., Marchand, A., Boyer, R., Préville, M., & Kairouz, S. (2010). A life-course and time perspective on the construct validity of psychological distress in women and men. Measurement invariance of the K6 across gender. *BMC Medical Research Methodology, 10*(1), 68.

Eno Louden, J., Skeem, J. L., & Blevins, A. (2013). Comparing the predictive utility of two screening tools for mental disorder among probationers. *Psychological Assessment, 25*(2), 405.

Esposito, C. L., & Clum, G. A. (2002). Social support and problem-solving as moderators of the relationship between childhood abuse and suicidality: Applications to a delinquent population. *Journal of Traumatic Stress, 15*(2), 137-146.

Evans Cuellar, A., McReynolds, L. S., & Wasserman, G. A. (2006). A cure for crime: Can mental health treatment diversion reduce crime among youth? *Journal of Policy Analysis and Management, 25*(1), 197-214.

Farrell, A. M., & Rudd, J. M. (2009). *Factor analysis and discriminant validity: A brief review of some practical issues*.

Fazel, S., Doll, H., & Långström, N. (2008). Mental disorders among adolescents in juvenile detention and correctional facilities: a systematic review and metaregression analysis of

25 surveys. *Journal of the American Academy of Child & Adolescent Psychiatry, 47*(9), 1010-1019.

Feehan, M., McGee, R., Raja, S. N., & Williams, S. M. (1994). DSM-III-R disorders in New Zealand 18-year-olds. *Australasian Psychiatry, 28*(1), 87-99.

Fergusson, D., Boden, J., & Hayne, H. (2011). Childhood conduct problems. *Improving the Transition*, 59.

Fergusson, D., Horwood, J., & Lynskey, M. T. (1997). Children And Adolescents. In P. M. Ellis & S. C. Collings (Eds.), *Mental health in New Zealand from a public health perspective* (pp. 136-163): Public Health Group, Ministry of Health.

Fergusson, D., Horwood, L. J., & Lynskey, M. T. (1993). Prevalence and comorbidity of DSM-III-R diagnoses in a birth cohort of 15 year olds. *Journal of the American Academy of Child & Adolescent Psychiatry, 32*(6), 1127-1134.

Fergusson, D., Poulton, R., Horwood, J., Milne, B., Swain-Campbell, N., & Baxter, J. (2003). Comorbidity and coincidence in the Christchurch and Dunedin longitudinal studies. *Wellington: Ministry of Social Development/Department of Labour/The Treasury*.

Fleming, T. M., Clark, T., Denny, S., Bullen, P., Crengle, S., Peiris-John, R., . . . Lucassen, M. (2014). Stability and change in the mental health of New Zealand secondary school students 2007–2012: Results from the national adolescent health surveys. *Australian and New Zealand journal of psychiatry, 48*(5), 472-480.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological methods, 9*(4), 466.

Forbes, D., Elhai, J. D., Miller, M. W., & Creamer, M. (2010). Internalizing and externalizing classes in posttraumatic stress disorder: A latent class analysis. *Journal of Traumatic Stress, 23*(3), 340-349.

Ford, J. D., Chapman, J. F., Pearson, G., Borum, R., & Wolpaw, J. M. (2008). Psychometric status and clinical utility of the MAYSI-2 with girls and boys in juvenile detention. *Journal of Psychopathology and Behavioral Assessment, 30*(2), 87-99.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, 39-50.

Forns, M., Abad, J., & Kirchner, T. (2011). Internalizing and Externalizing Problems. In R. J. R. Levesque (Ed.), *Encyclopedia of Adolescence* (pp. 1464-1469). New York, NY: Springer New York.

Fortune, S., Watson, P., Robinson, E., Fleming, T., Merry, S., Denny, S. (2010). *Youth'07: The health and wellbeing of secondary school students in New Zealand: Suicide behaviours and mental health in 2001 and 2007.* . Auckland.

Foster, M., Qaseem, A., & Connor, T. (2004). Can better mental health services reduce the risk of juvenile justice system involvement? *American Journal of Public Health, 94*(5), 859-865.

Foster, S., & Cone, J. D. (1995). Validity issues in clinical assessment. *Psychological Assessment, 7*(3), 248.

Gefen, D., & Straub, D. (2005). A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example. *Communications of the Association for Information systems, 16*(1), 5.

Ghanizadeh, A., Nouri, S., & Nabi, S. (2012). Psychiatric problems and suicidal behaviour in incarcerated adolescents in the Islamic Republic of Iran. *Eastern Mediterranean Health Journal, 18*(4), 311.

Gilbert, A. L., Grande, T. L., Hallman, J., & Underwood, L. A. (2014). Screening incarcerated juveniles using the MAYSI-2. *Journal of Correctional Health Care, 21*(1), 1078345814557788.

Goldstein, N. E., Arnold, D. H., Weil, J., Mesiarik, C. M., Peuschold, D., Grisso, T., & Osman, D. (2003). Comorbid symptom patterns in female juvenile offenders. *International journal of law and psychiatry, 26*(5), 565-582.

Goodwin, C. J. (2009). *Research in psychology: Methods and design*: John Wiley & Sons.

Graham, J. M., Guthrie, A. C., & Thompson, B. (2003). Consequences of not interpreting structure coefficients in published CFA research: A reminder. *Structural Equation Modeling, 10*(1), 142-153.

Green, J. G., Gruber, M. J., Sampson, N. A., Zaslavsky, A. M., & Kessler, R. C. (2010). Improving the K6 short scale to predict serious emotional disturbance in adolescents in the USA. *International journal of methods in psychiatric research, 19*(S1), 23-35.

Gretton, H. M., & Clift, R. J. (2011). The mental health needs of incarcerated youth in British Columbia, Canada. *International journal of law and psychiatry, 34*(2), 109-115.

Grisso, T., & Barnum, R. (2006). *Massachusetts Youth Screening Instrument, Version 2: MAYSI-2: User's Manual and Technical Report*: Professional Resource Press.

Grisso, T., Barnum, R., Fletcher, K. E., Cauffman, E., & Peuschold, D. (2001). Massachusetts Youth Screening Instrument for mental health needs of juvenile justice youths. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*(5), 541-548.

Grisso, T., Fusco, S., Paiva-Salisbury, M., Perrauot, R., Williams, V., & Barnum, R. (2012). The Massachusetts youth screening instrument-version 2 (MAYSI-2): comprehensive research review. *University of Massachusetts Medical School, Worcester*.

Grisso, T., & Quinlan, J. C. (2005). Massachusetts Youth Screening Instrument-Version 2. *Mental health screening and assessment in juvenile justice*, 99-111.

Grisso, T., Vincent, G., & Seagrave, D. (2005). *Mental health screening and assessment in juvenile justice*: Guilford Press.

Group, A. H. R. (2013). The health and wellbeing of New Zealand secondary school students in 2012. *Youth'12 Prevalence Tables*.

Hair, J. F., Anderson, R. E., Babin, B. J., & Black, W. C. (2010). *Multivariate data analysis: A global perspective* (Vol. 7): Pearson Upper Saddle River, NJ.

Hamerlynck, S. M., Doreleijers, T. A., Vermeiren, R., Jansen, L. M., & Cohen-Kettenis, P. T. (2008). Aggression and psychopathology in detained adolescent females. *Psychiatry Research, 159*(1), 77-85.

Harzke, A. J., Baillargeon, J., Baillargeon, G., Henry, J., Olvera, R. L., Torrealday, O., . . . Parikh, R. (2012). Prevalence of psychiatric disorders in the Texas Juvenile Correctional System. *Journal of Correctional Health Care, 18*(2), 143-157.

Hayes, L. (2009). *Juvenile suicide in confinement: A national survey*: DIANE Publishing.

Hayes, M., McReynolds, L. S., & Wasserman, G. A. (2005). Paper and Voice MAYSI-2 format comparability and concordance with the voice DISC-IV. *Assessment, 12*(4), 395-403.

Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*(3), 238.

He, J., & van de Vijver, F. (2012). Bias and equivalence in cross-cultural research. *Online readings in psychology and culture, 2*(2), 8.

Health, M. o. (2015). *Suicide Facts: Deaths and intentional self-harm hospitalisations 2012*. Wellington.

Health, M. o. (2016). *2015/16 New Zealand Health Survey* Wellington.

Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: a cautionary note on the usefulness of cutoff values of fit indices. *Psychological methods, 16*(3), 319.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115-135.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). Applied statistics for the behavioral sciences.

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal, 6*(1), 1-55.

Indig, D., Claudia Vecchiato, L. H., Rodney Beilby,, Julie Carter, U. C., Claire Gaskin, Eric Heller, Shalin Kumar,, & Natalie Mamone, P. M., Paul van den Dolder, and Gilbert Whitton. (2011). *2009 NSW Young People in Custody Health Survey: Full Report*: Justice Health.

Jackson, D. L., Gillaspy Jr, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: an overview and some recommendations. *Psychological methods, 14*(1), 6.

Jaggers, J., Young, S. R., McKinney, R., Bolland, K., & Church, W. T. (2013). Utilization of the Massachusetts Youth Screening Instrument-2 (MAYSI-2) with a Southern, African American Adolescent Male Population. *Journal of Forensic Social Work, 3*(1), 3-15.

James, L. M., Strom, T. Q., & Leskela, J. (2014). Risk-taking behaviors and impulsivity among veterans with and without PTSD and mild TBI. *Military medicine, 179*(4), 357-363.

Jones, J. A., & Waller, N. G. (2015). The normal-theory and asymptotic distribution-free (ADF) covariance matrix of standardized regression coefficients: theoretical extensions and finite sample behavior. *psychometrika, 80*(2), 365-378.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *psychometrika, 36*(2), 109-133.

Justice, M. o. (2016). *Trends in Child and Youth Prosecutions*.

Kaplan, R., & Saccuzzo, D. (2012). *Psychological testing: Principles, applications, and issues*: Cengage Learning.

Kerig, P. K., Moeddel, M. A., & Becker, S. P. (2011). *Assessing the sensitivity and specificity of the MAYSI-2 for detecting trauma among youth in juvenile detention.* Paper presented at the Child & Youth Care Forum.

Kersten, P., Dudley, M., Nayar, S., Elder, H., Robertson, H., Tauroa, R., & McPherson, K. M. (2016). Cross-cultural acceptability and utility of the strengths and difficulties questionnaire: views of families. *BMC psychiatry, 16*(1), 347.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S.-L., . . . Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological medicine, 32*(06), 959-976.

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., . . . Walters, E. E. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry, 60*(2), 184-189.

Kieling, C., Baker-Henningham, H., Belfer, M., Conti, G., Ertem, I., Omigbodun, O., . . . Rahman, A. (2011). Child and adolescent mental health worldwide: evidence for action. *The Lancet, 378*(9801), 1515-1525.

Kirmayer, L. J. (2001). Cultural variations in the clinical presentation of depression and anxiety: implications for diagnosis and treatment. *Journal of Clinical Psychiatry, 62*, 22-30.

Kovacs, M., Paulauskas, S., Gatsonis, C., & Richards, C. (1988). Depressive disorders in childhood III. A longitudinal study of comorbidity with and risk for conduct disorders. *Journal of affective disorders, 15*(3), 205-217.

Krynen, A., Osborne, D., Duck, I. M., Houkamau, C., & Sibley, C. G. (2013). Measuring psychological distress in New Zealand: Item response properties and demographic

differences in the Kessler-6 screening measure. *New Zealand Journal of Psychology, 42*(1), 69-83.

Kubiak, S. P., Beeble, M., & Bybee, D. (2012). Comparing the validity of the K6 when assessing depression, anxiety, and PTSD among male and female jail detainees. *International journal of offender therapy and comparative criminology, 56*(8), 1220-1238.

Lader, D., Singleton, N., & Meltzer, H. (2003). Psychiatric morbidity among young offenders in England and Wales. *International Review of Psychiatry, 15*(1-2), 144-147.

Leenarts, L., Dölitzsch, C., Schmeck, K., Fegert, J., Grisso, T., & Schmid, M. (2016). Relationship between Massachusetts Youth Screening Instrument-second version and psychiatric disorders in youths in welfare and juvenile justice institutions in Switzerland. *BMC psychiatry, 16*(1), 340.

Lennox, C., O'Malley, K., Bell, V., Shaw, J., & Dolan, M. (2014). The validity and clinical utility of the Massachusetts Youth Screening Instrument–version 2 (MAYSI-2) in the UK. *Criminal Behaviour and Mental Health*.

Lewinsohn, P. M., Rohde, P., & Seeley, J. R. (1996). Adolescent suicidal ideation and attempts: Prevalence, risk factors, and clinical implications. *Clinical Psychology: Science and Practice, 3*(1), 25-46.

Lynch, N. (2014). 'Girls Behaving Badly?'Young Female Violence in New Zealand. *Victoria University of Wellington Law Review: Special Issue in honour of Harriette Vine, First Woman Law Graduate, 45*(3), 509-524.

McClelland, G. M., Elkington, K. S., Teplin, L. A., & Abram, K. M. (2004). Multiple substance use disorders in juvenile detainees. *Journal of the American Academy of Child & Adolescent Psychiatry, 43*(10), 1215-1224.

McCoy, H. (2010). Using cognitive interviewing to explore causes for racial differences on the MAYSI-2. *Crime & Delinquency*, 0011128710388922.

McCoy, H. (2011). A path analysis of factors influencing racial differences on the Massachusetts Youth Screening Instrument–Version 2. *Journal of Offender Rehabilitation, 50*(3), 119-141.

McCoy, H. (2015). *Think Outside of the Maysi-2 Box and See Loneliness As a Red Flag.* Paper presented at the Society for Social Work and Research 19th Annual Conference: The Social and Behavioral Importance of Increased Longevity.

McCoy, H., Vaughn, M. G., Maynard, B. R., & Salas-Wright, C. P. (2014). Caution or Warning? A Validity Study of the MAYSI-2 with Juvenile Offenders. *Behavioral sciences & the law, 32*(4), 508-526.

McGee, R., Feehan, M., Williams, S., Partridge, F., Silva, P. A., & Kelly, J. (1990). DSM-III disorders in a large sample of adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 29*(4), 611-619.

McHugh, T., Forbes, D., Bates, G., Hopwood, M., & Creamer, M. (2012). Anger in PTSD: is there a need for a concept of PTSD-related posttraumatic anger? *Clinical psychology review, 32*(2), 93-104.

Mewton, L., Kessler, R. C., Slade, T., Hobbs, M. J., Brownhill, L., Birrell, L., . . . Chapman, C. (2016). The psychometric properties of the Kessler Psychological Distress Scale (K6) in a general population sample of adolescents. *Psychol Assess, 28*(10), 1232-1242.

Miller, L., McIntire, S. A., & Lovler, R. L. (2011). *Foundations of psychological testing: A practical approach*: Sage.

Miller, M., & Resick, P. A. (2007). Internalizing and externalizing subtypes in female sexual assault survivors: Implications for the understanding of complex PTSD. *Behavior therapy, 38*(1), 58-71.

Morgan, L., Scourfield, J., Williams, D., Jasper, A., & Lewis, G. (2003). The Aberfan disaster: 33-year follow-up of survivors. *The British Journal of Psychiatry, 182*(6), 532-536.

Newsom, J. (2012). Some clarifications and recommendations on fit indices. *USP, 655*, 123-133.

Nordness, P. D., Grummert, M., Banks, D., Schindler, M. L., Moss, M. M., Gallagher, K., & Epstein, M. H. (2002). Screening the mental health needs of youths in juvenile detention. *Juvenile and Family Court Journal, 53*(2), 43-50.

Nunally, J. C., & Bernstein, I. H. (1978). Psychometric theory. In: New York: McGraw-Hill.

Oakley-Browne, M., Wells, J. E., Scott, K. M., Kessler, R., & Üstün, T. (2008). *Te rau hinengaro: the New Zealand mental health survey*: World Health Organization.

Organization, W. H. (1992). *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*: Geneva: World Health Organization.

Organization, W. H. (2000). The World health report: 2000: Health systems: improving performance.

Otto, R. K., Greenstein, J., Johnson, M. K., & Friedman, R. (1992). Prevalence of mental disorders in the juvenile justice system. *In Cocozza, J.(ed.) 1992, Responding to the Mental Health Needs of Youth in the Juvenile Justice System.*

Paus, T., Keshavan, M., & Giedd, J. N. (2008). Why do many psychiatric disorders emerge during adolescence? *Nature Reviews Neuroscience, 9*(12), 947-957.

Peiper, N., Clayton, R., Wilson, R., & Illback, R. (2015). The performance of the K6 Scale in a large school sample. *Psychological Assessment, 27*(1), 228.

Penn, J. V., Esposito, C. L., Schaeffer, L. E., Fritz, G. K., & Spirito, A. (2003). Suicide attempts and self-mutilative behavior in a juvenile correctional facility. *Journal of the American Academy of Child & Adolescent Psychiatry, 42*(7), 762-769.

Peterson, R. A., & Kim, Y. (2013). On the relationship between coefficient alpha and composite reliability. In: American Psychological Association.

Plattner, B., Aebi, M., Steinhausen, H.-C., & Bessler, C. (2011). [Psychopathological and comorbid disorders of incarcerated adolescents in Austria]. *Zeitschrift fur Kinder-und Jugendpsychiatrie und Psychotherapie, 39*(4), 231-240; quiz 241-232.

Plattner, B., Kraemer, H. C., Williams, R. P., Bauer, S. M., Kindler, J., Feucht, M., . . . Steiner, H. (2007). Suicidality, psychopathology, and gender in incarcerated adolescents in Austria. *The Journal of clinical psychiatry*.

Pliszka, S. R., Sherman, J. O., Barrow, M. V., & Irick, S. (2000). Affective disorder in juvenile offenders: A preliminary study. *American Journal of Psychiatry, 157*(1), 130-132.

Polanczyk, G. V., Salum, G. A., Sugaya, L. S., Caye, A., & Rohde, L. A. (2015). Annual Research Review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry, 56*(3), 345-365.

Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research, 32*(4), 329-353.

Retz, W., Retz-Junginger, P., Hengesch, G., Schneider, M., Thome, J., Pajonk, F.-G., . . . Rösler, M. (2004). Psychometric and psychopathological characterization of young male prison inmates with and without attention deficit/hyperactivity disorder. *European archives of psychiatry and clinical neuroscience, 254*(4), 201-208.

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *psychometrika, 74*(1), 145-154.

Richards, I. (1996). Psychiatric disorder among adolescents in custody. *Australian and New Zealand journal of psychiatry, 30*(6), 788-793.

Richardson, R., Trépel, D., Perry, A., Ali, S., Duffy, S., Gabe, R., . . . Manea, L. (2015). Screening for psychological and mental health difficulties in young people who offend: a systematic review and decision model.

Roberts, R. E., Attkisson, C. C., & Rosenblatt, A. (1998). Prevalence of psychopathology among children and adolescents. *American Journal of Psychiatry, 155*(6), 715-725.

Robertson, A. A., Dill, P. L., Husain, J., & Undesser, C. (2004). Prevalence of mental illness and substance abuse disorders among incarcerated juvenile offenders in Mississippi. *Child Psychiatry and Human Development, 35*(1), 55-74.

Rohde, P., Seeley, J. R., & Mace, D. E. (1997). Correlates of suicidal behavior in a juvenile detention population. *Suicide and Life-Threatening Behavior, 27*(2), 164-175.

Rosan, A., Frick, P. J., Gottlieb, K. A., & Fasicaru, L. (2015). Callous-Unemotional Traits and Anxiety in a Sample of Detained Adolescents in Romania. *Journal of Evidence-Based Psychotherapies, 15*(1), 79.

Ruchkin, V. V., Schwab-Stone, M., Koposov, R., Vermeiren, R., & Steiner, H. (2002). Violence exposure, posttraumatic stress, and personality in juvenile delinquents. *Journal of the American Academy of Child & Adolescent Psychiatry, 41*(3), 322-329.

Russell, J. D., Marsee, M. A., & Ryals Jr, J. S. (2017). Identifying mental health issues in detained youth: Testing the structure and invariance of the Massachusetts Youth Screening Inventory–Version 2 (MAYSI-2). *Psychological Assessment, 29*(6), 720.

Satterfield, J., Swanson, J., Schell, A., & Lee, F. (1994). Prediction of antisocial behavior in attention-deficit hyperactivity disorder boys from aggression/defiance scores. *Journal of the American Academy of Child & Adolescent Psychiatry, 33*(2), 185-190.

Savalei, V. (2008). Is the ML chi-square ever robust to nonnormality? A cautionary note with missing data. *Structural Equation Modeling, 15*(1), 1-22.

Sawyer, M. G., Guidolin, M., Schulz, K. L., McGinnes, B., Zubrick, S. R., & Baghurst, P. A. (2010). The mental health and wellbeing of adolescents on remand in Australia. *Australian and New Zealand journal of psychiatry*, *44*(6), 551-559.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online, 8*(2), 23-74.

Shelton, D. (2001). Emotional disorders in young offenders. *Journal of Nursing Scholarship, 33*(3), 259-263.

Shulman, E. P., Bechtold, J., Kelly, E. L., & Cauffman, E. (2016). Mental Health Screening in Juvenile Justice Settings Evaluating the Utility of the Massachusetts Youth Screening Instrument, Version 2. *Criminal Justice Policy Review*, 0887403416650250.

Simic, M., & Fombonne, E. (2001). Depressive conduct disorder: symptom patterns and correlates in referred children and adolescents. *Journal of affective disorders, 62*(3), 175-185.

Spohn, C., & Beichner, D. (2000). Is preferential treatment of female offenders a thing of the past? A multisite study of gender, race, and imprisonment. *Criminal Justice Policy Review, 11*(2), 149-184.

Squeglia, L., Jacobus, J., & Tapert, S. F. (2009). The influence of substance use on adolescent brain development. *Clinical EEG and neuroscience, 40*(1), 31-38.

Ståhlberg, O., Anckarsäter, H., & Nilsson, T. (2010). Mental health problems in youths committed to juvenile institutions: prevalences and treatment needs. *European child & adolescent psychiatry, 19*(12), 893-903.

Starkweather, J. (2012). Step out of the past: Stop using coefficient alpha; there are better ways to calculate reliability. *University of North Texas. Research and statistical support. Retrieved September, 23*, 2013.

Stathis, S., Doolan, I., Fleming, R., Heath, K., Arnett, A., & Cory, S. (2008). Use of the Massachusetts Youth Screening Instrument to assess mental health problems in young people within an Australian youth detention centre. *Journal of paediatrics and child health, 44*(7-8), 438-443.

Steffensmeier, D., & Demuth, S. (2006). Does gender modify the effects of race–ethnicity on criminal sanctioning? Sentences for male and female white, black, and Hispanic defendants. *Journal of Quantitative Criminology, 22*(3), 241-261.

Stokes, M. L., McCoy, K. P., Abram, K. M., Byck, G. R., & Teplin, L. A. (2015). Suicidal Ideation and Behavior in Youth in the Juvenile Justice System A Review of the Literature. *Journal of Correctional Health Care, 21*(3), 222-242.

Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual review of clinical psychology, 5*, 1.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education, 2*, 53.

Team, R. C. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013. In.

Teplin, L. A., Abram, K. M., McClelland, G. M., Dulcan, M. K., & Mericle, A. A. (2002). Psychiatric disorders in youth in juvenile detention. *Archives of General Psychiatry, 59*(12), 1133-1143.

Thompson, B. (1997). The importance of structure coefficients in structural equation modeling confirmatory factor analysis. *Educational and Psychological Measurement, 57*(1), 5-19.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*: American Psychological Association.

Timmons-Mitchell, J., Brown, C., Schulz, S. C., Webster, S. E., Underwood, L. A., & Semple, W. E. (1997). Comparing the mental health needs of female and male incarcerated juvenile delinquents. *Behavioral sciences & the law, 15*(2), 195-202.

Tseng, W.-T., Dörnyei, Z., & Schmitt, N. (2006). A new approach to assessing strategic learning: The case of self-regulation in vocabulary acquisition. *Applied Linguistics, 27*(1), 78-102.

Uestuen, B., & Kennedy, C. (2009). What is "functional impairment"? Disentangling disability from clinical significance. *World Psychiatry, 8*(2), 82-85.

Urbina, S. (2014). *Essentials of psychological testing*: John Wiley & Sons.

Van Prooijen, J.-W., & Van Der Kloot, W. A. (2001). Confirmatory analysis of exploratively obtained factor structures. *Educational and Psychological Measurement, 61*(5), 777-792.

Vermeiren, R. (2003). Psychopathology and delinquency in adolescents: a descriptive and developmental perspective. *Clinical psychology review, 23*(2), 277-318.

Vincent, G. M., Grisso, T., Terry, A., & Banks, S. (2008). Sex and race differences in mental health symptoms in juvenile justice: The MAYSI-2 national meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry, 47*(3), 282-290.

Volkow, N. D., Baler, R. D., Compton, W. M., & Weiss, S. R. (2014). Adverse health effects of marijuana use. *New England Journal of Medicine, 370*(23), 2219-2227.

Voorhees, C. M., Brady, M. K., Calantone, R., & Ramirez, E. (2016). Discriminant validity testing in marketing: an analysis, causes for concern, and proposed remedies. *Journal of the Academy of Marketing Science, 44*(1), 119-134.

Vreugdenhil, C., Doreleijers, T. A., Vermeiren, R., Wouters, L. F., & Van Den Brink, W. (2004). Psychiatric disorders in a representative sample of incarcerated boys in the

Netherlands. *Journal of the American Academy of Child & Adolescent Psychiatry, 43*(1), 97-104.

Wang, P. S., Gruber, M. J., Powers, R. E., Schoenbaum, M., Speier, A. H., Wells, K. B., & Kessler, R. C. (2007). Mental health service use among Hurricane Katrina survivors in the eight months after the disaster. *Psychiatric Services, 58*(11), 1403-1411.

Wasserman, D., Cheng, Q., & Jiang, G.-X. (2005). Global suicide rates among young people aged 15-19. *World Psychiatry, 4*(2), 114.

Wasserman, G., McReynolds, L. S., Lucas, C. P., Fisher, P., & Santos, L. (2002). The voice DISC-IV with incarcerated male youths: prevalence of disorder. *Journal of the American Academy of Child & Adolescent Psychiatry, 41*(3), 314-321.

Wasserman, G., McReynolds, L. S., Schwalbe, C. S., Keating, J. M., & Jones, S. A. (2010). Psychiatric disorder, comorbidity, and suicidal behavior in juvenile justice youth. *Criminal Justice and Behavior, 37*(12), 1361-1376.