Research to Validate the New Zealand Police Youth Offending Risk Screening Tool (YORST) Phase I:

Screening and Assessment of Young Offenders Risk of Recidivism: Literature Review

A report prepared by

Dr Elaine Mossman
of the Crime and Justice Research Centre
Victoria University of Wellington

2010

New Zealand
POLICE
Nga Pirihimana O Aotearoa

Acknowledgements

Contents

Tables

Executive summary

1   Introduction

This report responds to a request by the New Zealand Police to conduct a literature review of risk screening instruments for youth offenders, with a particular focus on the methods used to establish validity and reliability of these tools, including validation across cultures. The purpose of the literature review is to inform a larger project that aims to assess the validity, reliability and predictive capability of the New Zealand Police Youth Offending Risk Screening Tool (YORST).

Between December 2009 and February 2010, researchers identified and then reviewed relevant research. The review focused on literature that met the following criteria:

- recent research (within the last ten years) unless earlier seminal research was identified

- factors that specifically predict youth *re-offending* (from here on referred to as recidivism) rather than studies that researched which young people who become offenders (i.e. first time offenders, often referred to as juvenile delinquency). This focus was in line with the intended purpose of the YORST which is to identify the likelihood of recidivism in young people who have already been in contact with police

- secondary reviews or meta-analyses of international research, with primary research limited to New Zealand-based sources and/or other recent relevant research not included in the reviews.

2   Factors associated with youth re-offending

Risk screening/assessment instruments, such as the YORST, are based on factors that have been found to reliably predict young offenders' rate of recidivism. These factors may be one of the following:

- **static risk factors** that are historical in nature and cannot be changed through an intervention (e.g. criminal history)

- **dynamic risk factors** which can be potentially changed, for example anti-social attitudes, negative peer associations, truancy or abuse of alcohol or drugs. Dynamic factors that, if targeted through an intervention, have been shown to have a causal link with recidivism are referred to as **criminogenic needs**

- **protective factors** which are characteristics or conditions that interact with risk factors to moderate or reduce their influence (e.g. having pro-social peers or a positive home environment).

The literature on factors that predict recidivism in youth is extensive and was beyond the scope of this report to fully review. Instead we focused on research that has been carried out in New Zealand and reviews or meta-analyses of the other research.

In general, there is considerable agreement over which factors are associated with recidivism. Factors most consistently reported to have a statistical association with future re-offending include:

- **demographic variables:** such as being male and of an ethnic minority, also criminal history factors, such as early onset of offending, and the frequency and severity of prior offending[1]

- **individual characteristics:** such as whether a young person abuses drugs and alcohol, the existence of mental health problems and/or whether or not the individual associates with delinquent peers

- **education variables:** such as whether a young person has evidence of poor educational engagement (e.g. low achievement/failure at school, truancy, expelled and/or excluded from school)

- **family characteristics:** for example poor parental supervision, family history of criminal activity, dysfunctional family dynamics

- **community characteristics:** such as a young person living in a disadvantaged neighbourhood and having a lack of attachment to that community.

It is important to recognise that the factors listed above are those most frequently found by researchers to be predictors of recidivism and does not tell us which factors are the strongest predictors of recidivism.

Comparative predictive strength of risk factors

To assess the comparative importance of particular risk factors it is necessary look at each factor's predictive strength. However, comparative information has been provided in a recent meta-analysis that looked at the predictive strength of 30 individual risk factors based on the findings of 23 studies. Out of the 30 individual risk factors, 24 were found to significantly predict recidivism. The strongest predictors were:

- early involvement with the criminal justice system (e.g. age of first contact and age of first incarceration)

- non-severe pathology

- family problems

- conduct problems

- ineffective use of leisure time

- association with delinquent peers.

Unfortunately these findings are predominantly based on international literature and some factors have less relevance to the New Zealand situation (e.g. age of first incarceration). Hence, when considering the appropriate content of the YORST priority should be given to factors found to be significant risk factors for recidivism for New Zealand young offenders. Many factors are similar to those identified

---

[1] Belonging to an ethnic minority has frequently been found to be associated with risk of recidivism, however, researchers caution that it is inappropriate to consider this as a risk factor. This is because it has been shown to be a proxy for a series of risk factors that are associated with marginalised groups (such as poverty, truancy, low levels of educational achievement) and it is these factors rather than the ethnic background that are predictors of recidivism. Inclusion of ethnicity in risk assessments would only add further to other systemic bias that already exists.

overseas, predictive factors that had particular relevance to the New Zealand Youth Justice System, included early and frequent contact with Child, Youth and Family and police, and a young person's experiences of Family group Conferences.

Risk factors for different groups

Young offenders are not a homogenous group and a key concern in validating instruments such as the YORST is whether the instrument's predictive ability and capacity to screen for criminogenic needs is valid across different sub-groups of young offenders, in particular:

- male and female young offenders

- the different age groups, 10 to 13 years and 14 to 16 years

- youth offenders belonging to different ethnic groups. Māori youth are a particularly important considering their over-representation in the youth justice system.

There has been considerable debate surrounding the extent to which risk factors are common or specific to certain groups. The lack of agreement on this issue highlights the need for careful attention to assessing the validity of the YORST across gender, ethnicity and age bands.

3 Measuring risk of recidivism

With the increasing recognition of the importance of risk assessment in relation to young offenders, a number of youth risk screening/assessment instruments have been developed. These instruments can be classified in terms of the historical context in which they were developed, their method of assessment and in terms of their intended purpose:

- **first generation risk assessment:** are unstructured or clinical assessments of young offenders' risk of recidivism based on an individuals professional judgement

- **second generation risk assessment:** are assessments grounded in the statistical association between a risk screening/assessment instrument and repeat offending (i.e. actuarial approach). The purpose of these second generation instruments was limited to prediction and classification or risk

- **third generation risk assessment:** are also grounded in the statistical association between a risk screening/assessment instrument and repeat offending but they have a dual purpose. In addition to risk classification and prediction these instruments are also used to inform intervention planning through the assessment of a young offender's criminogenic needs. As a result, they tend to be lengthy, using a wide range of risk and protective factors and focussing on dynamic factors that can be altered through intervention.

Structured assessments are favoured by many for their ability to increase consistency, objectivity and equity of youth justice decision-making. There is also evidence that structured assessments are more accurate. Another approach to assessment utilised by some instruments is structured professional judgement (SPJ). This is where a specific set of empirically derived factors are reviewed by a clinician to guide their final appraisal on the level of risk of any individual. This approach is intended to improve subjective decision-making by adding structure and improve actuarial decision-making by adding some rater discretion.

Methods of calculating risk

Instruments that produce risk scores, generally use one of two methods to calculate the score:

i.   using computer assisted statistical modelling, where relative weights can be assigned to items according to their predictive strength

ii. the Burgess method, where typically non-weighted values for each of a series of factors are simply added to produce a risk score.

Analysis of which method is more effective in predicting risk of recidivism has found the simple Burgess method can be equivalent and, in some cases, superior to the more sophisticated computerised modelling.

Attributes of risk screening/assessment instruments

Risk screening/assessment instruments can vary according to a number of different attributes such as:

* sources of information required for completion

* decision point within the criminal justice system that the risk screening/assessment is aiming to inform

* sensitivity and specificity

* type and number of items.

Understanding the attributes of different type of instruments will be useful in the evaluation of those of the YORST.

Instruments vary according to the sources of information required for completion. This can include police records, court records, school records, ratings by informants such as parents, teachers, social workers, treatment providers or diagnosis by mental health professionals. An instrument must be able to feasibly be completed using available resources, and taking into consideration the skill levels and type of experience of those carrying out the assessment.

Risk screening/assessment instruments tend to be tied to a particular decision point within the criminal justice system, and different points are associated with different questions and assessment needs and will therefore vary in their content and format.

* **At first intake or contact with police** – is youth appropriate for diversion?

* **Prior to court if charged** – does youth need secure pre-trial detention?

* **At court** – what is appropriate placement, security level, treatment plans?

* **If incarcerated** – what is the risk of suicide and/or violence? What types of interventions are most appropriate?

* **On community re-entry** – what is an appropriate level monitoring and intervention?

Information requirements at different points may also be jurisdiction specific. For example, compared to other jurisdictions, New Zealand Police Youth Aid Officers have a greater degree of autonomy in deciding the appropriate level of action police take for any young offender they come into contact with. This means information on criminogenic needs as well as risk of recidivism is of relevance.

The number and type of factors included in an instrument will depend on its purpose. If statistical prediction of recidivism is all that is required a short actuarial tool, with predominantly static variables, is likely to be perfectly adequate. Instruments with as few as five-items have been shown to reliably predict recidivism. However, if the aim is to identify areas for intervention or to be able to measure change over time, then it will require a more detailed assessment instrument, with priority given to dynamic factors that can provide information on both risk and criminogenic need.

When tools are being used to classify youth as high or low risk offenders another important consideration is the relative sensitivity and specificity of the tool (i.e. its ability to minimise false negatives compared to false positives). This relates to a tool's associated threshold for classifying a young person as high risk and, therefore, in need of intensive intervention. If the threshold is too low, there will be a number of young offenders being diverted to treatment unnecessarily. If it is too high, then there are likely to be chronic serious recidivists that are not identified who will not receive the intervention they need to reduce the likelihood of further offending.

While there is growing recognition of the advantages of standardised risk screening/assessment, there are also some limitations. Some argue that structured risk screening/assessment can result in over-simplification of complex social problems. Others point out that that these can diminish the practitioner's ability of to exercise skilled and professional judgement. Finally, it is important to understand that risk classification is based on probability and so being categorised high risk does not make a young person's offending a certainty.

4   Evaluation of risk screening/assessment tools

The overall aim of risk screening/assessment instruments is to enable improved quality and outcomes of youth justice decision-making. However, if a tool is used inappropriately, or it is invalid, this can have a deleterious effect on decisions and outcomes.

Validity and reliability

The validity and reliability of an instrument can be established using a number of different approaches and tests.

To be valid, an instrument must be shown to have appropriate content to be able to measure what it is intending to measure (content validity), if there are sub-scales, there needs to be evidence that they measure distinct constructs (construct validity), and, perhaps most importantly, that the instrument accurately measures what it is intending to measure (criterion validity).

There are two types of criterion validity:

- **Concurrent validity** is established by demonstrating agreement between the tool in question and another risk screening/assessment tool with established validity. Instruments with established validity that are commonly used to test concurrent validity include the YLS/CMI, PCL:YV, and the SARVY. The two instruments are completed at the same time and the level of agreement between the validated test and the tool being developed assessed

- **Predictive** validity relates to an instrument's ability to predict a criterion variable (i.e. recidivism). A tool is typically developed retrospectively on a 'construction' sample. However, once developed, it is critical for predictive validity to be cross-validated, and tested prospectively on a different 'validation' sample of young offenders to that from which the tool was first developed. Other concerns in testing

predictive validity are selecting an appropriate criterion variable and how to account for confounding variables such as the impact of interventions on recidivism

Equally important as validity, is the reliability of the instrument, the extent to which it is consistently able to produce the same result across time (stability reliability) and across raters (inter-rater reliability). Most risk screening/assessment instruments are based on subjective ratings by practitioners so establishing inter-rater reliability is most critical. This is most commonly tested using a matched pairs approach, where independent assessments of the same offender are made by two practitioners and compared for their level of agreement.

If an instrument is found to have low inter-rater reliability this can mean one of two things, that:

i.   there is a problem with the instrument and it requires revision

ii. there is a problem with the way raters have been trained in the use of the tool or with accompanying instructions.

Other evaluation considerations

In addition to establishing the validity and reliability of the instrument it is also important to evaluate the policies, protocols and training that surround the tool and to carefully consider if the tool is appropriate for the particular context in which it is to be used. It is important to know that the tool is being implemented as intended, that it is user friendly and appropriate for the local setting. Regardless of its validity, if it is incorrectly implemented or inappropriate for the context it will not achieve valid or useful results.

Evaluation activities might include a review of documentation by researchers, together with consultation with practitioners on their experiences, and in particular any problems, in using the instrument. Field observations are also a useful way to assess implementation issues.

The appropriateness of the instrument to the local context must also be evaluated. Is it feasible to complete the tool given the resources available and the skill set of staff, and does it provide information that is relevant to the needs associated with the specific youth justice decision point?

5   Risk screening/assessment tools

In evaluating the YORST it is useful to compare its format and content to that of other risk screening/assessment instruments (e.g. factors included, length, format, validity and reliability).

In the past, there has been limited risk screening of young offenders in New Zealand. This prompted the development of the YORST and its predecessor the ARNI.

A review of other risk screening/assessment carried out in New Zealand revealed that, to date, there has been greater progress in the risk screening of adult offenders. The Department of Corrections has successfully integrated an actuarial model of risk prediction into their adult offender management system. It is a computerised generated score based on static factors which has shown to have good predictive validity. Factors used in the calculation of individual risk scores include:

* current age

* gender

- age at first conviction

- number and seriousness of convictions

- number and type of previous sentences

- amount of time spent in custody.

In relation to young offenders, progress towards an integrated system of risk screening has been slower. There have been two standardised risk screening/assessment tools developed (RSYO and YRS), but neither appear widely used and validation testing is incomplete. There has also been one attempt to develop an actuarial model for young offenders by researchers in Canterbury. However, despite promising preliminary results, this model appears to have received no further attention.

Other screening of young offenders in New Zealand is that carried out by Child, Youth and Family social workers. Child, Youth and Family have developed a package of screening and assessment tools which has been successfully implemented nationally. This includes:

- **Cage:** a screening tool for drug and alcohol use (asks about recent alcohol and drug use followed by a further four standard questions)

- **Kessler:** is a screening tool for psychological distress (6-items)

- **Suicide tool:** screens for risk of suicide (primarily clinical judgement but provides three prompts for social workers to ask young person)

- **The Suicide Risk Assessment and Suicide Risk Management Plan:** intended to support the social worker to investigate the level of suicide risk and describe the steps they will take to respond to that risk

- **Wellbeing assessment:** is used to assess the needs and strengths of a young person and their family. The tool is primarily designed for youth justice, but can and is used in Care and Protection. It covers the following domains:
  o pattern of offending
  o family/whānau environment
  o education/employment
  o physical wellbeing
  o emotional wellbeing
  o attitudes
  o social interactions and peer relationships
  o spiritual and cultural identity.

This screening and assessment package is currently under review with the aim of introducing tools with more established validity and reliability.

Risk screening/assessment tools used overseas

Standardised risk and needs assessment and/or screening of offenders is well recognised overseas as best practice and has resulted in the development of a number of standardised instruments. A recent review identified 28 risk screening/assessment tools for young offenders. Just over half of these (n=16) are brief actuarial tools, with the remaining being more detailed tools assessing both risk and needs

(n=12). The review summarised details of 12 instruments that appear to be more widely used and/or had reported established validity and reliability which were divided into two categories.

i. **Brief / screening instruments** with 20-items and/or take less than 30 minutes to complete, and require no clinical training to administer. Instruments reviewed included the SECAPS Recidivism Index, North Carolina Assessment of Risk (NCAR), Arizona Risk/Needs Assessment (ARNA), Antisocial Process Screening Device rating scale (APSD), Psychopathy Content Scale (PCS).

ii. **More detailed assessment instruments** those more comprehensive with over 20-items and/or take greater than 30 minutes to complete. These may require clinical training to administer and included the Youth Level of Service / Case Management Inventory (YLS/CMI), Australian Adaptation of YLS/CMI (YLS/CMI-AA), the ASSET, Psychopathy Checklist - Youth Version (PCL:YV), Structured Assessment of Violence Risk in Youth (SARVY), Secure Care Psychosocial Screening (SECAPS), Victorian Offending Needs Indicatory for Youth (VOINY).

Considering the proliferation of risk screening/assessment tools that have now been developed it is useful to consider what is known about their comparative effectiveness. A recent meta-analysis of 28 risk screening/assessment instruments has shown that overall these instruments are considered to have a medium performance in their ability to predict recidivism. Slightly lower, but not dissimilar, to that found with risk screening/assessment instruments for adults. No one instrument was found to be superior to others, but the highest effect sizes (ability to predict recidivism) were found for the PCL:YV, CAFAS, YLS/CMI, SARVY and the ASSET. The YLS/CMI has the strongest evidence of being valid across gender and indigenous and non-indigenous youth and across different jurisdictions.

Researchers concluded that if service planning is a priority then instruments like the YLS/CMI that measure criminogenic need are likely to be most useful. However, if brevity and efficiency are a higher priority then locally developed actuarial models should be considered.

6   Implications for the YORST

The key purpose of this literature review is to inform a larger programme of research which aims to assess the validity and reliability of the New Zealand Police YORST. It would therefore, be premature to make conclusive appraisals of the YORST at this stage. However, in light of the literature review the following points can be made in relation to the YORST.

- **Type:** the YORST is best described as a third generation instrument as it includes both static and dynamic factors and aims to inform case management in addition to estimating risk of recidivism. It development has been informed through a combination of practitioner input and review and consideration of empirically-based risk factors.

- **Length:** the YORST is longer than many of the brief second generation instruments that primarily aim to predict risk (several of these instruments accurately predict recidivism with five or fewer items). However, the YORST is considerably shorter than other third generation tools which, in addition to predicting risk, aim to assess criminogenic needs in order to inform case management.

- **Decision point:** the YORST has been designed to be used at the point when children (10-13 years) and youth (14-16 years) come to police notice (due to alleged offending). This represents a large group of young offenders, the majority of whom will be low risk offenders, however, completion criteria prioritises its use with those likely to be higher risk young offenders. The YORST must therefore be capable of differentiating level of risk among this particular group.

- **Information requirements:** the intended purpose of the YORST is to identify a young person's risk of re-offending and to provide the foundation for a targeted and appropriate response. It is also hoped the YORST will be able to measure change in level of risk, thus providing a measure of the impact of police interventions. The inclusion of dynamic factors in the YORST therefore fits with its stated objectives, however, careful attention needs to be given to whether the number of dynamic items on the YORST will sufficiently inform case management and detect change in risk over time.

- **Scoring:** the scoring system for the YORST appears to be based on the cumulative risk model or the Burgess method, where item responses are summed to produce a risk score.

- **Content:** all items on the YORST measure (or represent a proximal estimate) of risk factors identified by research as statistically significant predictors of recidivism. Some items identified in the literature that act as strong predictors of recidivism have not been included in the YORST (e.g. age at first commitment (placement in a correctional facility), non-severe pathology, conduct problems, effective use of leisure time, length of first incarceration, number of prior commitments). However, some of these factors are less relevant to the New Zealand youth justice system (e.g. incarceration of young offenders) and some factors may be considered inappropriate for police officers to assess.

Research to evaluate the YORST is in its early stages. This review has presented details of other established risk screening/assessment instruments which provide a useful basis to compare the format and content of the YORST (e.g. factors included, length, format, validity and reliability). Subsequent evaluation activities will include an analysis of the accuracy and completeness of YORST data being collected, followed by testing of its validity and reliability. This review has highlighted the need for careful consideration in assessing the validity of the YORST across gender, ethnicity and age bands. It may also be useful to monitor the uptake of the tool and evaluate the policies, protocols and training that impact on how the tool is implemented.

1   Introduction

This report responds to a request by the New Zealand Police to conduct a literature review of the process of risk screening of youth offenders and the methods and results of validation of risk screening tools, including validation across cultures. It has been prepared by researchers from Crime and Justice Research Centre (CJRC).

The purpose of the literature review is to inform a larger project that aims to assess the validity, reliability and predictive capability of the New Zealand Police Youth Offending Risk Screening Tool (YORST).

1.1   Background

A range of youth justice decisions are made on the basis of judgements around a young person's likelihood of re-offending together with their areas of criminogenic need. When offending by a young person first comes to the attention of the police, judgements on risk and needs inform decisions around what action to take, whether a custodial remand is appropriate, and what, if any, intervention or referrals should be made. A primary concern in making these decisions is public safety, how can the public be best protected from immediate risk of harm from young offenders, and also through reducing recidivism by providing appropriate interventions to youth. However, if these judgements on risk and need are made through informal and unsystematic procedures, this can lead to inconsistency and bias in the decision process. As a result, standardised risk and need assessment and/or screening of offenders is now well recognised as best practice (Andrews & Bonta, 1998; The Werry Centre, 2009; Vincent et al., 2009) and has consequently been implemented in many overseas jurisdictions.

The need for standardised assessment of young offenders in New Zealand has been highlighted in several government reports (Ministry of Justice and Ministry of Social Development, 2002ab Ministry of Justice, 2002) and also by New Zealand researchers (Maxwell et al., 2002). As a result, in 2003 New Zealand Police began work to develop a risk screening tool. An early version of the tool they developed was known as ARNI and its later version the YORST.

YORST is a 14 item questionnaire that is filled out by Police Youth Aid Officers on all child offenders (10-13 years) at their second offence and/or incident,[2] or who have been referred to Child, Youth and Family for care and protection due to serious offending; and all youth offenders (14-16 years) being referred for a Youth Justice Family Group Conference (a copy of the YORST appears in Appendix 1). The YORST has five intended outcomes (New Zealand Police, 2009) to:

- screen for levels of risk in young people
- screen for areas of risk in young people
- enable Child Youth and Family to ensure that further assessment and interventions are targeted for young people who offend
- enable the investigation of the impact of Police interventions on risk of re-offending
- indicate the likelihood (or risk) of the young person re-offending so that appropriate decisions can be made about the intervention pathways for individuals.

---

[2]   An incident is something relating to the child or young person's behaviour which is not an offence but has been reported to police and is recorded in the Police National Intelligence Application (NIA). For example incidents may include being reported as a missing person, truancy or being picked up late at night.

In order for the YORST to effectively carry out these objectives it needs to be both valid and reliable. A key purpose of this literature review is to summarise existing research on what is known about factors that predict recidivism and other validated risk screening/assessment instruments, as a preliminary step in the evaluation of the content and properties of the YORST. More rigorous testing of the validity and reliability of the YORST will follow on from the literature review.

1.1.1   Purpose of screening youth for risk of re-offending

Standardised instruments to screen and/or assess young offenders' risks and needs have been developed for a number of reasons. Perhaps the primary reason is to classify youth into groups which vary in their likelihood of repeat offending. In this way limited resources can be allocated to those most in need (i.e. high risk offenders) and where the greatest impact on reducing crime can be made. Information on level of risk can also be used to guide judicial action, (for example placement decisions such as whether remanding a young person in custody is appropriate).

The other main purpose behind standardised screening and/or assessment is to provide information on a young person's criminogenic needs such as whether they have a substance abuse problem or issues related to educational engagement. This allows effective case management and a targeted approach to treatment, where interventions are matched to areas of need, that if addressed have been shown to reduce the likelihood of recidivism. Screening tools can also be used to identify possible psychiatric problems in young offenders, for example identifying whether there is a risk of suicide that needs to be addressed when detaining youth in a custodial situation.

Standardised instruments are also used to increase consistency in the decision-making process. Instruments can act as a checklist to ensure each offender who passes through the system is considered according to a specific set of factors. This standardisation increases consistency between practitioners and also allows greater transparency in the decision-making process. Krysik and LeCroy (2002) also argue it makes decision-making more equitable as it ensures the same factors are considered for each youth.

The collection of standardised information also provides the opportunity to examine, analyse and compare data for large numbers of young offenders within and across regions (Kalb, 2006). This can assist in the targeting/allocation of resources and also serves to improve the evidence base on which youth justice policy decision are made.

1.1.2   Differentiating between 'screening' and 'assessment' instruments

The YORST has been designed as a risk 'screening' tool, and hence understanding the attributes and purpose of screening tool is central to this review. The concept of 'screening' is closely associated with 'assessment'; these related concepts are borrowed from the field of medicine, where they are clearly defined and understood. However, the fit and differentiation of 'screening' and 'assessment' in relation to risk of recidivism in young offenders is more problematic.

Using medical criteria, Bailey et al. (2006) defines screening as essentially a filtering system to identify individuals who require further assessment. This screening can be applied rapidly and inexpensively to a wide number of undiagnosed individuals. For example a general health screening may include a few questions around a person's emotional stability or risk of heart

_____

disease; those who screen positively would then be referred for further assessment to determine if they met agreed criteria for a clinical diagnosis of a mental health disorder (e.g. depression or depression). The assessment would require more specialist training, and typically involve data gathering from a range of sources such as a structured diagnostic interview or clinical appraisal. This increased level assessment would inevitably involve increased costs and time compared to the brief screening. Thus the combination of screening followed by assessment is a way to maximise limited financial resources.

Within youth justice, when 'screening' and 'assessment' are applied in relation to risk of recidivism their differentiation becomes less clear. This is because rather than predicting and then diagnosing the presence of an actual mental health disorder or physical illness, instruments are used to 'diagnose' future behaviour, that is, whether or not a young person will re-offend. A fuller assessment may increase the accuracy of the prediction, but cannot not verify its presence.

Many of the more comprehensive instruments that predict risk of recidivism, although typically referred to as risk 'assessment' instruments are, in effect, really just more detailed screening tools (Hoge, 2002). For example, the Youth Service Level: Youth Version (YSL:YV) provides an overall risk of recidivism score but also indicates likelihood of problems in other areas such as substance abuse which require further assessment. Similarly, individuals who scored above a certain cut-off with the Psychopathy Checklist; Youth Version (PCL:YV) would then be referred for further assessment and diagnosis (Murrie and Cornell, 2002). To further, complicate matters even very short instruments that use just a few variables to predict recidivism are also referred to as assessment instruments, for example the Arizona Risk Need Assessment (ARNA) which is based on only five variables.

Within the field of youth justice, rather than their ability to diagnose, level of detail and length of time to complete, appear more relevant in differentiating screening and/or assessment of risk of re-offending. Weatherburn et al. (2007) noted the large volume of young offenders coming to the attention of police and/or the court make extensive assessment impractical and prohibitively expensive. As a consequence Weatherburn et al. (2007) pointed to the need for screening rather than assessment. Vincent (2006) also described screening tools as those that were designed to be quick and able to be administered to every youth at entry or intake into some part of the youth justice system.

In addition to length and level of detail, screening and assessment for risk of re-offending appear to be differentiated from assessment through the types of decisions they influence. Weatherburn et al. (2007) suggest that the initial decision confronted by courts and juvenile justice agencies when dealing with young offenders making their first contact with the justice system "is not what forms of intervention to take place, but whether to intervene (in any substantial way) at all." These authors suggest screening should be seen as a form of 'triaging' using a few objective and readily obtained indicators of risk to identify those youth who ought to be referred for more careful assessment. Vincent (2006) also views screening tools as those assisting with short-term decision-making (e.g. whether a custodial remand was appropriate), while assessment tools were seen as longer and more comprehensive instruments designed to assist with case management and service planning (Vincent, 2006).

It is clear the differentiation between screening and assessment is less applicable in relation to instruments designed to predict risk of recidivism in young offenders. Therefore, for the purposes of this report we have chosen to combine the two terms 'screening' and 'assessment' when referring generically to such instruments i.e. 'screening/assessment instruments'. The terms will, however, be used individually if they are part of the specific title of an instrument or if an instrument is being referred to for the specific purpose of either screening or assessment.

1.2   Approach to reviewing the literature

Between December 2009 and February 2010, researchers identified and then reviewed all relevant research (details of the methods used to locate and review the literature appear in Appendix 2).

Research that identifies factors that are associated with, and can explain, youth offending comprises a substantial body of literature and was beyond the scope of this report to fully review. Consequently, this review focused on literature that met the following criteria:

- recent research (within the last ten years) unless earlier research was seminal work of relevance

- factors that specifically predicted youth *re-offending* (from here on referred to as recidivism) rather than studies that research which young people become offenders (i.e. first time offenders, often referred to as juvenile delinquency). This focus was in line with the intended purpose of the YORST which is to identify the likelihood of recidivism in young people who have already been in contact with police

- secondary reviews or meta-analyses of overseas research, with primary research included limited to that based on New Zealand samples and/or other recent relevant research not included in the reviews.

As noted above, the YORST has been designed to be a screening tool to identify youth at risk of recidivism. Hence another intended focus of the review was to be research that described and validated 'screening' tools for risk of recidivism in youth as opposed to 'assessment instruments'. As already alluded to above, this proved difficult. We were able to locate research that described and validated 'screening' tools that were designed to identify mental health concerns for young offenders (see Werry Centre, 2009 for a review), however, the vast majority of research on instruments that specifically predicted *recidivism* were described as 'risk assessment instruments. However, on closer examination of the literature, many of the instruments referred to as 'assessment instruments' were technically screening devices.' As a result, the review has included research that refers to both risk assessment and screening instruments.

1.3   Structure of the report

This report has been structured as follows:

- Chapter two summarises the key factors that have been found to be reliably associated with youth recidivism. These factors are important when evaluating the content of variables included in the YORST.

- Chapter three presents the main methods that are used to predict risk of recidivism in youth and introduces the different types and attributes of risk screening/assessment instruments that have evolved. Understanding the attributes of different type of instruments will be useful in considering those of the YORST.

_____

- Chapter four reviews important aspects in evaluating the appropriateness of risk screening/assessment tools, including how instruments are tested for their reliability and validity. This can be used to inform subsequent research to validate the YORST.

- Chapter five will describe youth risk screening/assessment tools that have been used in New Zealand and those from other jurisdictions which are accepted as having good validity and reliability. These instruments will be useful to compare the attributes of the YORST against, and will also assist in selecting a tool that is appropriate to test the concurrent validity of the YORST against.

- Chapter six will summarise the implications from findings of the literature review in relation to the YORST.

2 Factors associated with youth re-offending

It is generally accepted that a small but significant number of serious young offenders are responsible for a disproportionate amount of offending (McLaren, 2000; Moffit, 1993; Wilson & Rolleston, 2004; Youth Justice Board, 2005). According to the Principle Youth Court Judge Andrew Becroft the same is true in New Zealand, with the majority of youth offending (50-75%) being carried out by a group of just 5 -15% of young offenders, who then go on to continue to offend as adults (Becroft, 2004). This knowledge, that a relatively small percentage of offenders are responsible for a disproportionately large number of offences, makes identification of these persistent young recidivists a critical element of any effective crime prevention strategy.

The issue facing police and other youth justice professionals is to determine which of the youth they come into contact with, are likely to go on to become these serious and persistent offenders. If identified, resources can then be directed to those most in need and where the greatest impact on reducing crime can be made.

As McLaren (2000) states, apart from their early and repeated offending, the main characteristics of persistent young recidivists are the number of problems they experience.

> Together with their families, persistent young offenders show a range of problems that may include substance abuse, criminal behaviour, accommodation difficulties, poverty, unemployment, mental health problems, violence, neglect and abuse of every type imaginable, poor education, and more (McLaren, 2000, p.9).

This chapter takes a closer look at factors that have been found to reliably predict this group's high risk of recidivism. The developments of any risk screening/assessment instruments, such as the YORST, are based on these factors. Before presenting the relevant research the chapter will begin with some definitions of the different types of factors together with other key concepts that underpin this area of research.

2.1 Defining the terms

Thompson and Putnins (2003) state that the idea of 'risk' is used in two related ways when thinking about juvenile crime. Firstly, risk refers to the overall likelihood that a young person will engage in criminal behaviour (i.e. the risk of offending or re-offending). Secondly, 'risk' may be used to describe the specific conditions that are associated with offending (i.e. the risk factors). This latter use of the term is further separated into static or dynamic risk factors:

- **Static risk factors** are historical in nature, and are ones that cannot be changed through intervention. For example, demographics, history of conduct disorder or age of first offence are all static risk factors for recidivism (Hoge, 2002). Vincent et al. (2009) describe static risk factors as 'risk markers' which are associated with but do not cause recidivism.

- **Dynamic risk factors** are those that can potentially be changed, for example anti-social attitudes, negative peer associations, truancy or abuse of alcohol or drugs (Cottle et al., 2001, Hoge, 2002). Dynamic factors that if targeted through intervention have been shown to have a causal link with recidivism are referred to as **criminogenic needs** (Schwalbe, 2009).

A key limitation of static factors is that because they are not amenable to change, unlike dynamic factors, they cannot be used to detect any changes in risk level. If a youth is classified as 'high risk' using static factors, re-assessment of risk post-intervention, will remain the same.

> **It is important to understand that although the presence of a particular risk factor may increase the probability of re-offending this does not make offending a certainty (Shader, 2003, p.2).**

The presence of a risk factor will impact differently on different young people – for example, people living in the same neighbourhood can respond differently to their surroundings. This can partly be explained by what are referred to as protective factors.

Protective factors have been viewed as the absence of risk, but also as characteristics or conditions that interact with risk factors to moderate or reduce their influence (Shader, 2003; Youth Justice Board, 2005). Commonly identified protective factors for recidivism include pro-social peers and a positive home environment (Hoge, 2002). Protective factors are closely related to the concept of resilience, the presence of which can explain why some youth experiencing multiple risk factors desist from crime. Schwable (2009) argues that protective factors can be utilised in a number of ways to reduce risk, these can:

- reduce exposure

- disrupt effects

- act as a buffer

- open compensatory opportunities.

The roles and importance of the different types of risk factors vary depending on whether the primary goal is identification and classification of a young person's level of risk or if it is the case management and rehabilitation of a young person. Often both these goals are important, their differing roles are demonstrated in the risk-need-responsivity model presented below.

2.1.1   Risk, need and responsivity

Understanding the differing goals and purposes of the different factors are central to understanding and evaluating different screening/assessment instruments for young offenders. The risk-need-responsivity model of offender assessment and rehabilitation (Andrews, Bonta & Hoge, 1990) has become a hugely influential approach to working with offenders both overseas and here in New Zealand. The differing role of static, dynamic and responsivity factors are clearly highlighted through this model. The three components of this model are:

i)   **The *risk principle* asserts that criminal behaviour can be reliably predicted and that treatment should focus on the higher risk offenders.**

The *risk principle* is based on the premise that higher levels of service should be reserved for higher risk cases, because they respond better to intensive service than to less

intensive. There is also concern the placement of lower risk youth into higher risk programming may actually increase their likelihood of offending (Gavazzi et al., 2008).

Static factors, particularly demographics and historical offence details such as age of first arrest tend to be the strongest predictors of recidivism, and hence have taken precedence in identifying and classifying young offenders into different levels of risk (Upperton & Thompson, 2007; Vincent, 2006). Although dynamic factors such as family processes, and school related issues have been identified as having significant association with high risk offenders, debate continues about the usefulness of such information in predicting recidivism beyond what already is accounted for by static factors (Gavazzi et al., 2008; Hoge, 2002; Thompson, 2005).

However, a limitation of static factors in relation to young offenders is that, compared to adults, their criminal history is shorter, so there is less information (e.g. frequency of offending, number of times incarcerated) on which to base predictions (Grace et al., 2006; Wilson & Rolleston, 2004).

ii) **The *need principle* highlights the importance of identifying criminogenic needs and targeting them in treatment.**

Dynamic factors play an important role in relation to the *need principle*. Here only factors that are dynamic (i.e. amenable to change) are relevant, more specifically, those that have been shown to have a causal link with recidivism (i.e. criminogenic needs). Static factors have no role in relation to the need principle.

iii) **The *responsivity principle* suggests interventions should take into account treatment and client factors that influence success.**

The *responsivity principle* relates to a different set of factors. These factors are not necessarily related to criminal activity but are relevant to the way youth react to different types of intervention (e.g. reading ability, learning styles, self-esteem, and motivation for treatment). Consideration of culture or ethnic group membership becomes relevant when assessing the responsivity of particular youth to programmes (Hoge, 2002).

The use inclusion of different types of factors in different types of instruments is revisited in Chapter three.

2.2 Identifying predictors of recidivism

The literature on factors that predict recidivism in youth is extensive and it was beyond the scope of this report to fully review. Instead we focused on research that had been carried out in New Zealand, and reviews or meta-analyses of the other research.

Despite the extensive research in this area, as we reviewed this material it became clear that making authoritative statements about the risk factors for recidivism was not as straight forward as perhaps first thought. Further, presenting an over-simplified summary of risk factors without appropriate cautions would be misleading (O'Mahony, 2009; Webster et al., 2006).

There were a number of key characteristics of research, which can vary significantly, and impact greatly on the interpretation and implications of their findings.

_____

- **Offending vs re-offending:** Factors that are predictive for first time offenders tend to be based on broad samples of youth prior to their identification as offenders. It is not feasible to make meaningful assumptions about predictors of recidivism based on this research (Cottle et al., 2001). Factors predictive of recidivism must be derived from research on those individuals who have already come into contact with the youth justice system. These latter studies are the focus of this review.

- **Measure of recidivism:** Research studies use 'recidivism' differently and according to the focus of a particular project. For example recidivism has been based on self-reported criminal behaviour or official statistics such as re-arrest by police or a further prosecution (i.e. reconviction). Recidivism has also been measured by frequency of offending, seriousness, or time at large. Each variation can impact on the type and strength of predictive factors. Another factor is the length of follow-up which varied from six months to over five years and often longer in longitudinal studies. It is also important to note that recidivism based on official statistics relates only to offending where there had been a prosecution (i.e. a measure of reconviction not re-offending).

- **Type of offending:** Some studies focussed on any re-offending, others focused on violent or serious re-offending only.

- **Measurement of variables:** Commonly reported predictor variables often vary considerably in how they are measured. For example, substance abuse was sometimes based on lifetime incidence, and on other occasions based on current use; sometimes it was related to any use of substances and other times more specifically substance abuse or dependency. The ability to identify factors is also limited by how accurately they are recorded. The ability to assess the influence of substance use will depend on the accuracy of a young offender's self-reported use, which is recognised as problematic.

- **Type of offender:** Research is often based on samples of predominantly Caucasian male young offenders, yet as Cottle et al. (2001) makes clear, offenders are not a homogenous group, and the applicability of findings to sub-groups of offenders was not always considered.

Despite these factors having significant impacts on the applicability and relevance of findings, many studies failed to appropriately operationalise key variables.[3] Given the limitations listed above, it is clear caution is needed when interpreting collated findings from this field of research.

In Table 2.1 we have listed all the empirically-based factors associated with recidivism identified in the literature (all factors listed have been demonstrated to have a statistical association with future re-offending). We have grouped factors under six domains (demographic factors, criminal history factors, individual factors, school factors, family factors, community factors). Individual factors listed are those identified in either New Zealand-based reviews of literature (McLaren, 2000; Crawford & Kennedy, 2009); overseas reviews and meta-analyses (Cottle et al., 2001, Hawkins, et al., 200; Shader, 2003; Thompson & Putnins, 2003; Watt, et al., 2004; Youth Justice Board, 2008); or have been identified through New Zealand longitudinal research and/or other New Zealand research (Galletly, 2006; Grace et al., 2006; Maxwell & Morris, 1999; Maxwell et al., 2004).

---

3 This included some very comprehensive studies, such as the Youth Justice Board for England and Wales report on Risk and Protective Factors (2005), which did not make a clear distinction between initial offending and recidivism when discussing risk factors.

Table 2.1:   Predictive factors for recidivism identified in literature

Demographic factors
• Sex/gender • Current age • Race/culture/ethnicity
Criminal history factors
• Age at first offence • Number of prior offences • Type/severity of offence
Individual factors
• Alcohol/drug substance abuse • Friendships with anti-social peers (including gang involvement) • Early violent/anti-social behaviour • Attitudes that condone offending and drug misuse • Running away from home • Ineffective use of leisure time • Low intelligence, neurodevelopment or cognitive impairment • Mental health problems (including aggression and anger, hyperactivity and impulsivity, conduct disorder) • Victim of physical/sexual/emotional abuse • Poor relationships with peers • Alienation and lack of social commitment/accountability • Referred to CYF for care and protection or Youth Justice
School factors
• Low engagement with school/truancy • Expelled or excluded from school • Low achievement/failure at school • Stealing at school • Bullying • Frequent changes of school • Special education/identified as learning disabled
Family factors
• Poor parental supervision/low involvement • Poor family functioning/family conflict/violence • Low income/poverty • Poor/ inconsistent/harsh discipline • Family history of criminal activity • Foster/state care / disrupted family bonds • Parental attitudes that condone anti-social and criminal behaviour • Poor relationship with parents • Frequent changes of home/transience • Parents who abuse alcohol or drugs • Young mother/birth problems
Community factors
• Community disorganisation and neglect • Living in disadvantaged neighbourhoods/poverty • Availability of drugs • Lack of neighbourhood attachment

Factors from the **demographic** and **criminal history domains** were significant static risk factors identified in many of studies, in particular being male, and early onset of offending (e.g. Cottle et al 2001; Crawford & Kennedy, 2008; Fergusson & Horward, 2002; Galettly, 2006; Johansson & Kemf-Leonard, 2009; McLaren, 2000; Miller & Lin 2007; Thompson & Puinin, 2003; Youth Justice Board, 2008). Within the **individual factors** domain the more commonly identified factors included alcohol/drug substance abuse (e.g. Stoolmiller & Blechman, 2005); also friendships with anti-social peers, mental health problems and early violent/anti-social behaviour.

Risk factors in the **school domain** that were consistently identified across studies were low engagement with school/truancy, expelled or excluded from school and low achievement/failure at school (e.g. Arnull et al., 2005; Crawford & Kennedy, 2008; Hawkins et al., 2000; Maxwell et al., 2004; Miller & Lin, 2007; Moffitt & Caspi, 2001).

In the **family domain,** the most frequently identified risk factors included poor parental supervision/low involvement with the young person (e.g. Cottle et al., 2001; Fergusson & Horward, 2002; Maxwell & Morris,1999; Maxwell et al., 2004; Moffitt & Caspi, 2001; Shader, 2003); also poor family functioning/family conflict/violence; physical/sexual/emotional abuse; and a family history of criminal activity (Arnull et al., 2005). Across the studies reviewed, **community factors** were considered less significant overall than those of individual, school and family domains.

It is important to note, however, that the frequency with which a factor had been identified as significantly predictive of recidivism, is as much a reflection of which factors had been included in different studies than as an objective measure of their level of importance. Some research has only studied the association of static factors (e.g. Galletley 2006), others have focused on dynamic (e.g. Gavassi et al., 2007), and many others have used both. Table 1 provides a fairly comprehensive list of the range of possible factors, but it lacks details on their comparative importance. This is provided by assessing their predictive strength, which is reviewed in the next section (2.3).

Long and short-term risk factors: Another consideration in reviewing the above factors is whether these are long or short-term factors. Van der Laan et al (2009) identify long term risk factors such as impulsivity, lack of social skills, inadequate parental supervision, and poor school performance and argue that serious youth recidivism has been explained by an accumulation of these long-term risk factors and a lack of protective factors. However, they argue that short-term risk factors are situationally relevant factors that increase the likelihood that an individual will commit an offence. In their study, short-term risk factors included the presence of co-offenders, the absence of guardians and being under the influence of alcohol or drugs prior to committing an offence.

Van der Laan et al (2009) found that an accumulation of long-term risk factors had a stronger association with serious recidivism than the short-term risks. However, they did find that short-term risks or situational factors, such as an absence of guardians or having used alcohol were also important, and they concluded that serious youth recidivism could be the "result of an accumulation of long term risk factors in a number of domains, the presence of short-term risk factors, and their interaction".

2.3   Predictive ability of different factors

A more useful analysis of the utility and predictive ability of the different factors is provided by a recent meta-analysis carried out by Cottle and colleagues (2001). The predictive strength of 30 individual risk factors is presented. This information is relevant to the development of risk screening/assessment tools as ideally items included in the instrument will be those most strongly correlated with recidivism.

This meta-analysis is the most comprehensive to date of international literature on predictors of recidivism in young offenders and includes 23 studies, conducted between 1983 and 2000. These studies represent 15,265 young offenders with a mean age of 14.7 years, 83% of whom were male and 48% Caucasian.

Thirty individual predictors of recidivism were divided into eight domains (demographic information, offence history, family and social factors, educational factors, intellectual and achievement scores, clinical factors and formal risk assessment).[4]   Raw statistics from each study were converted to correlation coefficients and were normalised using Fisher's transformation formula (see Cottle et al., for details of formula). The effect sizes were then used to calculate an overall weighted effect size for each variable (Zr). The mean levels of significance were then calculated.

Table 2.2 shows the individual risk factors ranked by weighted mean effect size (i.e. predictive strength). All but six factors significantly predicted recidivism. The strongest predictors were age at first commitment, age at first contact with the law, a history of non-severe pathology and family problems. Unfortunately the study is based predominantly on US research and it was difficult to see what the New Zealand equivalent of 'age of first commitment'. According to the studies that assess age of first commitment, it appeared to be age of first detention in a correctional facility. This is difficult to translate to New Zealand, as offending by children under the age 14 years, unless very serious, is dealt with as a care and protection matter through Child Youth and Family rather than as a youth justice matter. Only those older than 14 years will be placed in a youth justice residential facility.

In terms of the eight domains, Cottle et al. (2001) found offence history and family and social factors were consistently associated with recidivism. Other domains contained individual variables that were significant but were less consistent across the whole domain.

The major difference across the studies reviewed, and that of Cottle et al. (2001) was in the schooling domain. Previous reviews had found school factors to be a significant predictor, but the results of the meta-analysis found very low correlations between recidivism and predictors such as school attendance (-.048) and achievement (-.028).

---

[4]   Age was not included as a demographic variable because the meta-analysis involved only juveniles – typically ranging from 14-18 years of age.

Table 2.2:   Predictive strength of factors associated with recidivism in young offenders

Variable	Zr	N	k
1. Age at first commitment	-.346**	720	3
2. Age at first contact with the law	-.341**	1,225	8
3. Non-severe pathology	.305**	953	7
4. Family problems	.277**	1,054	5
5. Conduct problems	.255**	1,667	7
6. Effective use of leisure time	-.233**	588	2
7. Delinquent peers	.204**	1,525	7
8. Length of first incarceration	.187**	641	3
9. Number of out-of-home placements	.184**	424	2
10. Number of prior commitments	.174**	585	3
11. Type of crime	.159**	10,267	7
12. Standardised achievement score	-.153**	506	3
13. Substance abuse	.149**	1,111	6
14. Full scale IQ score	.-142**	1,756	5
15. History of special education	.130*	432	2
16. Risk assessment instruments	.118**	10,353	6
17. History of abuse	.112**	9,949	5
18. Gender (male)	.111**	9,761	3
19. Verbal IQ score	-.111*	716	4
20. Single parent	.070**	10,501	5
21. Severe pathology	.069	346	2
22. Race (minority)	.067**	10,121	6
23. Socio-economic status	.065**	10,363	3
24. Number of prior arrests	.058**	10,155	7
25. School attendance	-.048	299	2
26. Parent pathology	.047	529	3
27. Performance IQ score	-.31	491	2
28. School report of treatment	-.028	10,025	6
29. History of treatment	.019	9,366	2
30. Substance use	.014	9,366	2
Note: Zr = weighted mean effect size; k = number of unique samples; *p< .01 **p< .001			

2.3.1 Ethnicity as a predictor variable

Considerable debate has centred on whether or not ethnicity and/or racial classifications are an adequate and/or appropriate predictor of offending and/or recidivism (Cunneen, 2005; Gottfredson & Snyder, 2005). Those who support the use of ethnicity as a predictor argue that apprehension, arrest and conviction statistics provide a realistic portrayal of offending. As such, an ethnic group's over representation in criminal justice statistics is regarded as an accurate reflection of that group's rate of offending. In contrast, those who caution against the use of ethnicity as a predictive variable stress that:

- apprehension and conviction rates are a reflection of a systemic bias within the criminal justice system

- the use of ethnicity and/or racial classifications to predict offending and/or recidivism have been a misunderstanding of correlation versus causative relationships. An apparent relationship between marginalised ethnic groups and rates of criminal offending have been erroneously interpreted as causative.

Systemic bias

A number of studies have examined New Zealand arrest and conviction processes and have found evidence of systemic bias that place Māori at higher likelihood of police contact and conviction than non-Māori. For instance, Fergusson, Horwood & Lynskey (1993) examined rates of police contact amongst young people up to 14 years of age and found that young Māori were 2.9 times more likely to have contact with the police than non-Māori. Similar results were found for slightly older Māori youth who were found to be three times more likely to be apprehended, prosecuted and convicted than young non-Māori (Te Puni Kokiri, 2000). Further, Māori with a history of offending were more likely to be convicted than a non-Māori with the same offending history and social background (Fergusson, 2003) and Maxwell et al. (2004) found that young Māori were more likely to appear in the youth court than young non-Māori and were therefore more likely to be exposed to a range of more severe outcomes than those dealt with at a lower level. Similar findings have been reported in the United Kingdom and the United States where marginalised ethnic groups are more likely to be over represented in the criminal justice system (Rutter, Giller and Hagell, 1998).

Correlations between ethnicity and/or racial classification and recidivism

While correlations between ethnic identity and criminal offending have been identified, it is inappropriate to use such associations to infer causation (Gottfredson & Snyder, 2005). Such misconceptions have furthered an erroneous belief that those identified as belonging to marginalised ethnic categories are more likely to engage in criminal offending and are more likely to re-offend. Rather than ethnicity and/or race per se these classifications act as a proxy for a series of factors commonly associated with marginalised groups (such as poverty, truancy, low levels of educational achievement).

There are now many who highlight the inappropriateness of using ethnicity and/or racial classification as a predictive variable of recidivism (Morris & Maxwell, 1999; Gottfredson & Snyder, 2005). To assess and then direct services based on ethnicity and/or racial classification, will only add to other systemic bias that already exists.

2.4   Factors predictive of recidivism in New Zealand young offenders

It is important to consider risk factors that specifically predict recidivism in New Zealand youth. A recent study by Miller and Lin (2009) suggested generic or 'off the shelf' tools are less predictive than a locally developed risk screening/assessment tool. Applicable research is limited to the longitudinal research carried out in Dunedin and Christchurch, some individual research studies on groups of young offenders (e.g. Maxwell & Morris, 1999; Maxwell et al., 2004), and more recently an attempt to develop an actuarial model to predict recidivism in New Zealand youth (Galletly, 2006, Grace et al., 2006).

2.4.1   Adolescent onset/adolescent-limited and early onset/life course persistent

We are fortunate in New Zealand to have two longitudinal studies, both of which have produced valuable findings in relation to New Zealand young offenders. This type of research collects data from the same individuals over many years and thus has access to a greater source of information on potential risk factors than is possible with prospective research. In particular rich data are collected on early childhood experiences and psychosocial factors.

Using these data researchers from the Dunedin Multidisciplinary Health and Development study have made an internationally significant contribution to understanding youth offending through the identification and testing of two distinct groups of young offenders (Moffit, 1993; Moffit & Caspi, 2001' Odgers et al., 2008):

   i)   early onset/life-course persistent offenders (LPC)
   ii)   adolescence onset/adolescent–limited (AL) offenders.

LCP offenders make up only 5% to 6% of juvenile offenders, but commit crime at every life stage and are responsible for the majority of crimes committed by juvenile offenders (Moffit, 1993). Hence, identification and early intervention with this group is essential (Crawford & Kennedy, 2008). Crime committed by the AL group, as indicated by their label, is limited primarily to adolescence (Moffit, 1993), although, this does not mean that the crime they commit is not significant. A recent Ministry of Justice Report (Crawford & Kennedy, 2008) suggests that almost half of serious youth violence is attributable to offenders who show few signs of difficult behaviour before they enter their teens (i.e. AL offenders). The authors suggested that interventions for AL offenders need to target more proximal risk factors such as early adolescent offending, disengagement from school, having antisocial peers and substance abuse. Clearly, identifying these two groups of offenders and then applying appropriate interventions at the appropriate stage is critical to address youth crime.

These two groups can be distinguished by different predictor variables. The distinguishing factors of the LCP offender are the early childhood onset of antisocial behaviour, coupled with inadequate parenting during childhood and neurocognitive problems, and temperament and behavioural problems (Moffitt & Caspi, 2001). It is thought the persistent criminal behaviour is a result of inherited (genetic) or acquired neurodevelopmental deficits that interact with the environment during childhood development (Moffit, 2006, cited in Crawford & Kennedy, 2008). Such pathological backgrounds are not present in the childhoods of the AL offenders and because their pre-delinquent development is healthy, Moffit and Caspi (2001) suggest most young people who become AL are able to desist from crime when they become more mature.

Moffitt and Caspi (2001) found that both LCP and AL youth say that they have delinquent peers. However, when prior behavioural history is controlled, peers-related delinquency no longer predicts the delinquency of LCP offenders but continues to predict the delinquency of AL offenders. The authors found that while LCPs *attract* delinquent peers during adolescence, ALs are *attracted to* and influenced by delinquent peers (Moffitt and Caspi, 2001).

Research from the Christchurch Health and Development Study questioned however, whether different aetiological factors identified the two groups (Fergusson et al., 2000, Fergusson & Horwood, 2002). Fergusson et al. (2000) found a common set of aetiological factors that acted cumulatively to predict group membership. This supported earlier research by the same team that found multiple social, economic and family disadvantages in childhood predict later antisocial behaviour (Fergusson et al., 1994).

Recent debate has also suggested the dual classification of young offenders may be over simplified. Research by Fergusson et al. (2000) identified four groups, 'non-offenders', 'moderate risk offenders', 'adolescent-limited' and 'chronic offenders / life-course persistent'. A more recent study of the Dunedin cohort (Odgers et al., 2008) also proposed a modification to the original typology. The category of AL offenders was not as robust as originally thought, with researchers finding that many of the AL males continued to offend beyond adolescence. Odgers et al. (2008) redefined these young people as 'adolescent onset' rather than 'adolescent limited'.

Understanding the different types of young offender and how they might be identified plays a critical role in the development of youth justice policy. However, as Fergusson and Horwood (2002) point out, much of the information used to identify the LCP or AL offenders comes from the collection of early childhood data and for most children and young people coming to notice for offending, this information is not available. Distinguishing between the two groups is difficult when the available information is restricted to current behaviour and family background and functioning.

2.4.2   Risk factor research on New Zealand young offenders

Gabrielle Maxwell, Allison Morris and colleagues have previously carried out research studies on New Zealand on young offenders who had attended a Family Group Conference (FGC). Maxwell and Morris (1999) studied a sample of 108, while a later study by Maxwell et al. (2004) studied a large sample of 1003 young offenders. While this research was focussed primarily on the impact the FGC has on the lives of the young people, but both studies also carried out some analysis on factors that predicted recidivism.

Adverse early life experiences: The above research, in general, supported the findings of the New Zealand longitudinal research and other studies showing that adverse early life experiences particularly related to family background, together with early onset of anti-social behaviour predicted persistent re-offending (Maxwell and Morris, 1999; Maxwell et al., 2004). Related to educational background, Maxwell et al. (2004) found lack of school qualifications was also a powerful predictor of recidivism (pp 213-214).

Child welfare and criminal history factors: Also similar to overseas research, referral to child welfare agencies and criminal history factors were found to predict recidivism in these New Zealand youth. Maxwell et al. (2004) found that previous referral to CYF (for child protection or youth justice matters) was associated with recidivism. Interestingly, they found that the

seriousness of the offence was unrelated to recidivism, but that the number of offences and the number of types of offences were important. In this New Zealand sample, the number of different types of offences was the factor most strongly associated with recidivism.

Specific to the New Zealand context, both studies found outcomes and a young persons' experience of the FGC impacted on subsequent re-offending. Young people with negative experiences of FGCs, particularly where bad outcomes had resulted in them feeling ashamed, were more likely to re-offend (Maxwell and Morris, 1999; Maxwell et al., 2004).

Demographics: Maxwell et al. (2004) found sex and ethnicity were associated with recidivism – males were nearly three times more likely to re-offend than females. They also found that although there was no significant difference between Māori and New Zealand European re-offending, Pacific young people were only half as likely to re-offend as the rest. They suggested that ethnicity per se was likely to be less important in predicting re-offending than other variables associated with ethnicity. Another key finding from Maxwell and Morris's (1999) study was cultural pride and knowledge of one's ancestry acted as a protective factor for young offenders (Maxwell & Morris, 1999).

Using multivariate analysis, Maxwell et al. (2004) found five factors that significantly predicted recidivism:

- boys were three times more likely to re-offend than girls
- Pacific young people, were only half as likely to re-off3end as others
- a previous history of care and protection (1.16 times more likely)
- previous referrals for youth justice matters (1.27 times more likely)
- previous youth court appearances (1.63 times more likely).

Further analysis that focused on early life events found that a lack of school qualifications emerged as the most important predictor of recidivism, along with a history of involvement with Child Youth and Family.

2.4.3   Actuarial models based on New Zealand young offenders

An actuarial model to predict recidivism in New Zealand young offenders has been the subject of a Canterbury University Master's thesis written by Anna Galletly (2006). The model she developed was then included as part of the evaluation of the Reducing Youth Offending Programme (RYOP) carried out by Grace et al. (2006).

Galletly (2006) developed three actuarial models to predict different measures of recidivism in a sample of 500 young offenders. The measures of recidivism were:

M1 – referral to Child Youth and Family (CYF) for a youth justice matter

M2 – prosecution for any re-offending

M3 – prosecution for serious re-offending. Variables for inclusion in the model were a range of static factors from CYF's CYRAS database, and Police National Intelligence Application (NIA) database (see Appendix 3 for full list of variables considered)

Interestingly there were slight differences in CYRAS variables that were predictive for different CYF regions. Galletly (2006) suggested differences were due to demographic variability with a higher proportion of Māori and Pacific youth in Auckland. There were also regional variations in

relation to age at first intake, with Christchurch having earlier and more frequent CYF contact compared to other regions.

In addition to differences in predictor variables, there were also regional differences in the outcome measure, re-offending. Re-offending was highest in Christchurch (60%), followed by Auckland (58%) and a lower 50% over the rest of country. These variations in predictor variables and outcomes measures resulted in the first model (M1) being a hybrid model being developed that allowed for these differences across regions.

The most successful model was the second (M2) that predicted *prosecution for any re-offending*. Significant variables that were used in the model were a number of demographic and historical variables from CYRAS combined with two variables from NIA data on frequency of police contact.

Out of 27 variables accessed from CYRAS data the following six were significant predictors and included in the M2 as a hybrid model that allowed for the regional differences:

- ethnicity
- age at first intake
- age at first youth justice intake
- gender
- number of prior care and protection referrals
- number of prior court dates.

Out of six variables accessed from Police National Intelligence Application (NIA) data the following two were significant predictors and included in the model:

- number of prior intelligences
- number of prior occurrences.

This model had a moderately good accuracy on the development sample (AUC=.73) and also continued to perform at well on a second validation sample of a further 500 young offenders (AUC=.738). See section 3.6 for interpretation of AUC statistics.

2.5   Risk factors for different groups

Young offenders are not a homogenous group (Cottle et al., 2001) and a key concern in validating the YORST is whether its predictive ability and capacity to screen for criminogenic needs is valid for different sub-groups of young offenders in particular:

- male and female young offenders

- the different age groups 10 to 14 years and 14 to 16 years

- youth offenders of different ethnicity. Māori youth are a particularly important group considering their over-representation in the youth justice system.

Understanding whether the risk factors that predict recidivism vary for different groups is a useful way to consider whether or not one tool can be used across these groups.

_____

2.5.1   Gender

While being male has repeatedly been found to be a significant predictor of recidivism, there are mixed views on whether there are different risk factors that predict recidivism in young women and men. Researchers, using data from the New Zealand longitudinal studies, generally conclude that there is little or no difference in risk factors for young men and women (Fergusson & Horward, 2002; Johansson and Kempf-Leonard 2009; Olders et al, 2008). Other researchers have suggested that research that incorporates a greater range of factors, including those more relevant to females can find differences (Emeka & Sorenson, 2009); Holtfreter and Cupp, 2007).

According to Holtfreter and Cupp (2007), commonly used risk screening/assessment instruments fail to capture variables specific to females, such as victimisation and sexual/physical abuse. They suggest female delinquency is more likely to result from family problems, traumatic events, and mental health issues and histories characterised by physical and/or sexual abuse (see Emeka and Sorenson, 2009 for a review). However, a recent study by Johansson and Kempf-Leonard (2009) tested a theory of specific female risk factors for recidivism but failed to find a direct relationship between abuse and more serious, violent and chronic offending.

Another study conducted by Emeka and Sorenson (2009) involving 1647 under 15 year olds on probation in Texas found that females differed from their male counterparts on the main risk factors associated with female re-offending. Factors that predicted recidivism in females over a two year period included truancy, running away, a history of child abuse, trauma and victimisation.

The debate is set to continue on whether there are gender specific risk factors for recidivism and hence whether there should be different tools for males and females. However, a recent meta-analysis of risk screening/assessment instruments by Schwalbe (2008) found the predictive validity of a number of existing risk screening/assessment instruments was equivalent for males and females.

2.5.2   Ethnicity and/or racial classification

As discussed earlier, the inclusion of ethnicity as a risk factor is very contentious, however, there is clear acceptance that assessment processes must be applicable across all relevant ethnic groups. Unfortunately, despite a clear need, there is little literature on how risk assessment processes need to differ for different groups (Werry Centre, 2009). In New Zealand, longitudinal studies have found once factors associated with being Māori are taken into account, similar risk factors predict offending for Māori and non-Māori youth (Fergusson, 2003). However, other New Zealand studies have found a number of ethnic specific protective factors to be significant. For instance:

- a sense of strong cultural identity (Morris & Maxwell, 1999)
- supportive whānau (Roguski, 2009).

Overseas analysis of the applicability of some risk screening/assessment instruments for different ethnic groups have found mixed results. Some researchers have found the predictive

ability of instruments to be stable across different ethnic groups (e.g. ASSET,[5] Baker et al., 2005; Arizona Risk Needs Assessment (ARNA), Krysik & LeCroy, 2002) whereas others have found different factors did need to be taken into account for different ethnic groups (Schwalbe, 2006). It will be important to consider and test the validity of the YORST across different ethnic groups.

2.5.3   Age

The YORST has been designed for use with child offenders aged 10-13 years and young offenders aged 14-16 years. Most research on recidivism that has informed the development of the YORST is based on young offenders in the latter age group, a key concern is whether the YORST is also valid for this younger group of child offenders, or if they require an age specific instrument.

Mulvey (2005), cited in Vincent et al. (2009) argues that evaluating risk requires consideration of the developmental stage of the youth and the social context in which the behaviour occurs. This is because different risk indicators at different ages can mean different things and so levels of risk can change over time. Unfortunately, locating research that identified which risk factors for repeat offending specifically in children was difficult.

There is considerable research on the different risk factors that predict first time offending for different age ranges (i.e. the prediction of early-onset child offenders or late onset adolescent-limited). This research informs us that an early-onset of offending is predictive of recidivism in adolescents and into adulthood, however, tells us little about how to differentiate which child offenders will continue; yet research shows considerable variation in the duration and seriousness of subsequent offending in children. For example, Vincent et al. (2009) pointed out that at least 50% of children who initiate pervasive and serious antisocial behaviour between the ages of 6 and 12 do not develop into seriously antisocial adults.

In New Zealand research that focused on child offenders (Maxwell and Robertson, 1995) provided a clear profile of those children who went on to become serious and/or persistent child offenders. They were: victims of inadequate care or abuse, difficult to manage, absconded, performed poorly in school, experimented with alcohol and substances and many came from families where criminal behaviour and substance abuse are common.

Vincent et al. (2009) highlighted that behaviours which may pose a significant risk for child offenders (e.g. smoking before the age of 12), for adolescents may be seen as part of the normal developmental cycle of experimentation (Vincent et al., 2009). Others have found that, parental influence is more important in the early years in identifying young recidivists, where as sibling influence (i.e. delinquent siblings) was more important later in the teenage years (Farrington, 1997, cited in Arnull et al., 2005). Domburgh et al. (2009) noted that, in general, recidivism in child offenders is characterised by high levels of individual deviancy, social disadvantage and the lack of positive social bonds.[6] However, they go on to caution that

---

[5]   ASSET is not an acronym but the name given to the assessment tool.

[6]   This varied slightly between different levels of persistence: a high level of individual deviance was found predictive of serious persistence (n=95) whereas this was combined with social disadvantage in the moderately persistence group (n=117).

persistence and seriousness of offending are two different dimensions of an offending career and both should be taken into account when considering early onset re-offending careers.

The lack of research makes it difficult to conclude if child offenders are likely to require an age specific screening tool, but the limited research available suggests careful consideration will need to be given to the applicability of the YORST to child and young offenders.

2.6   Summary

This chapter has summarised the key factors (both static and dynamic) that have been found to be reliably associated with youth recidivism. In general there was considerable agreement over which factors these are. Static factors include demographic variables such as being male and of an ethnic minority and criminal history factors such as early onset of offending, and the frequency and severity of prior offending. While ethnic minority status has frequently been found to be associated with risk of recidivism, researchers caution that it is inappropriate to consider this as a risk factor as it had been shown to be a proxy for a series of risk factors that are associated with marginalised groups (such as poverty, truancy, low levels of educational achievement). Inclusion of ethnicity in risk assessments would only further other systemic bias that already exists.

Other factors were more dynamic in nature, included individual characteristics such as whether they abused drugs and alcohol, had mental health problems and/or associated with delinquent peers. Engagement problems with schooling were also consistently identified as risk factors for recidivism. There were also a number of family-related risk factors (e.g. poor parental supervision, family history of criminal activity and other dysfunctional family characteristics). Some studies also identified characteristics of the community a young person lived in as increasing their risk of re-offending.

It was noted, however, that the frequency with which a factor had been identified as significantly predictive of recidivism, is as much a reflection of which factors had been included in different studies than as an objective measure of their level of importance. The comparative importance of particular risk factors is determined by assessing their predictive strength. The findings of a recent meta-analysis of 23 studies provided this comparative information.

Early involvement with the criminal justice system was found to be the strongest predictor of recidivism (age of first contact and age of first incarceration). Other strong predictors included non-severe pathology, family problems, conduct problems, ineffective use of leisure time and association with delinquent peers.

These factors are important when evaluating the content of variables included in the YORST. However, review findings are based predominantly on international literature and some factors have less relevance to the New Zealand situation (e.g. age of first incarceration). Hence, when considering appropriate content of the YORST priority should be given to factors found to be significant risk factors for recidivism for New Zealand young offenders. It will also be important to consider whether risk factors included in the YORST are applicable to different sub-groups of young offenders (e.g. female, younger offenders, and offenders of different ethnicity).

3 Measuring risk of recidivism

Having reviewed the various factors that have been found to predict recidivism, we now turn to the methods used to measure risk. This chapter will briefly summarise the different ways judgements about a young person's risk of re-offending have traditionally been made. This will be followed by a description of the types of instruments that have evolved and the sources of information required for such assessments to be made. Issues around when to measure risk will be discussed, followed by a discussion of some of the attributes that are considered important in risk screening instruments. The chapter will end with an acknowledgement of some of the limitations of risk screening/assessment instruments.

3.1 Methods of assessing a young person's risk of recidivism

There are essentially three methods that are used to make judgements about a young person's level of risk in relation to recidivism.

- **Unstructured assessments:** also referred to as clinical assessments are based on an individual's professional judgement rather than assessment against a specific set of risk factors.

- **Actuarial decision-making:** this is a structured statistical approach to assessing risk, where risk scores are produced based on an empirically derived model of a set of factors found to be associated with recidivism. Models are typically developed using statistical analysis of large samples of offenders.

- **Structured professional judgement (SPJ):** a specific set of empirically derived factors are reviewed by a clinician to guide their final appraisal on the level of risk of any individual. This approach is intended to improve subjective decision-making by adding structure and improve actuarial decision-making by adding some rater discretion.

Structured assessment (either actuarial models or SPJ) are favoured by many. This relates to concerns that subjective decision-making used in 'unstructured assessments' can result in inconsistency, and has the potential for judgements to be influenced by the philosophies and different levels of experience and knowledge of those making the assessment (Kalb, 2006). Krysik and LeCroy (2002) suggest that structural risk prediction brings more equity and objectivity into youth justice decision-making because the same factors are considered for each youth.

There is also evidence that judgements based on structured assessments are more accurate (Baumann et al., 2005; Dawes et al., 1989; Grove & Meehl, 1996; Mills, 2005). A meta-analysis of 130 studies of human health and behaviour, that included criminal justice issues, showed actuarial techniques were on average 10% more accurate than unstructured clinical prediction (Grove et al., 2000: cited in Kalb 2006). This was supported by another meta-analysis that focused on adult risk measures that found actuarial methods far superior to unstructured professional judgement in predicting recidivism (Hanson & Morton-boughton, 2009). However, while actuarial methods are favoured by many, Vincent (2006) notes that their typical reliance on static factors limits their clinical utility. Models based on static factors cannot measure change in risk, hence, once a youth is classified as 'high risk'– this will not change post-intervention if re-assessed.

Comparisons made between the two types of structured assessments, actuarial and structured professional judgements found that structured professional judgements had improved predictive validity over actuarial-based decisions assessments (Douglas et al., 2005, cited in Vincent et al., 2009).

Curiously, research on young offenders in Australia found risk predictions from a structured inventory (Youth Level of Service/Case Management Inventory, YLS/CMI) were not significantly better than unstructured subjective ratings of youth justice professionals using a 0 to 10 point scale, with both performing well (Upperton & Thompson, 2007).

3.1.1 Predicted probability or cumulative risk model

Actuarial modelling identifies the risk factors that are the best predictors of recidivism. Once risk factors have been identified and models have been developed, there are two main methods for producing estimates of risk of recidivism. These can be calculated with computer assistance to enable appropriate weight of relevant variables (according to their predictive strength).[7] Variables can also be transformed to maximise their predictive ability (e.g. log transformation). Such methods can produce individualised probability scores for any individual (e.g. New Zealand Corrections Department ROC*ROI).

The other method is for the model and risk factors to be translated into a simple index that can be completed and scored by hand. The latter method is referred to as the Cumulative Risk Model or the Burgess method, where typically non-weighted values for each of a series of factors are simply added to produce a risk score.

Interestingly, analysis of which method is more effective in predicting risk of recidivism has found the simple Burgess method is equivalent and, in some cases, more superior to the more sophisticated computerised modelling (Gottfredson & Snyder, 2005; Kalb, 2006; Krysik & LeCroy, 2002). One explanation is that the simpler Burgess method is less susceptible to 'over-fitting' of models than computerised models. The Burgess method can also be more popular with practitioners, as the contribution of different items to the overall score is transparent and easily understood.

3.2 Types of instruments

The limitations of unstructured assessments have resulted in the development of a range of structured instruments. A useful classification of these young offender risk screening/assessment instruments is provided by Schwalbe (2007) based on the work by Andrews et al. (2006). Instruments are described in a historical context and are closely associated to the development methods described above, but are further differentiated by their intended purpose. Schwalbe describes three generations of screening/assessment instruments:

---

[7] The application of this method and the use of weightings was able to slightly improve the predictive accuracy of scores produced by the UK's ASSET (Baker et al., 2002; 2005).

_____

First generation risk assessment: are essentially the unstructured assessments of young offenders levels of risk used to classify young offenders by their perceived level of risk. Schwalbe defines them as the "impressionistic assessments of individual juvenile justice professionals without the aid of structural devices (Schwalbe, 2007, p.450.)"

Second generation risk assessment: are those assessments grounded in the statistical association between a risk screening/assessment instrument and repeat offending (i.e. actuarial approach). The purpose of these second generation instruments was limited to prediction and classification or risk. Schwalbe suggests these are often brief (i.e. less than 14-items) and are based on factors with the highest statistical association with recidivism; which tend to be static factors (e.g. offence history). These factors are then gathered together in some form of index.

Some second generation instruments, rather than being actuarially developed, are adaptations of the 'Model Risk Assessment Instrument' method recommended by the Office of Juvenile Justice and Delinquency (OJJDP), for example the nine-item, North Carolina Assessment of Risk (NCAR) instrument (Howell, 1995). According to Schwalbe (2007) these instruments look similar to actuarially-developed tools because they are brief and have similar items but they do not have the same statistical rigor.

Third generation risk assessment: are also grounded in the statistical association between a risk screening/assessment instrument and repeat offending but they have a dual purpose. In addition to risk classification and prediction they are also used to inform intervention planning through the assessment of a young offender's criminogenic needs. As a result, they tend to be longer, using a wider range of risk and protective factors and tend to focus on dynamic factors that can be altered through intervention (YJB, 2008). They tend to be closely associated with psychometric type instruments used in psychology, using multi-item measurement scales, often with several sub-scales, and assume an underlying factor structure. In addition to an overall risk score, they can measure constructs like personality or impairments. Many of the more widely used risk instruments fall into this category (e.g. YSL/CMI, PCL:YV; ASSET).

The ASSET is the main risk screening/assessment instrument used in the UK. It is best described as a third generation instrument as it assesses both risk and needs. In addition to information on the statistical association of different factors with risk, its development was also informed by the use of a specialist advisory panel comprised of representatives from a range of agencies and by feedback from practitioners obtained by piloting draft versions (Youth Justice Board, 2008).

In relation to adult offenders, Andrews et al. (2006) have added a **fourth generation** instrument – in addition to measuring risk and need this instrument also aims to capture more about responsivity factors that are relevant to the way an offender will potentially respond to an intervention (e.g. reading ability, learning styles, self-esteem, motivation for treatment, and culture and ethnicity factors).

Hanson and Morton-Bourgon (2009) suggest the debate over how instruments should be 'structured' is ongoing and is not only about the relative predictive accuracy of specific tools, but also the purpose of the risk assessment and the role of professional judgment. Schwalbe (2007) points to mixed messages in the literature – where some research advises policy makers to develop jurisdiction specific instruments using an actuarial approach – based on evidence that risk screening/assessment instruments do not generalise well to other jurisdictions (e.g. Miller & Lin, 2007). While on the other hand, some 'off the shelf' instruments have been implemented and validated in multiple settings (including many of the third generation instruments).

3.3 Attributes of risk screening/assessment instruments

As detailed above, risk screening/assessment instruments can vary considerably based on their intended purpose and method of development. There are a number of other attributes that also differentiate instruments and impact on the appropriateness of a specific tool for a particular purpose. These include the:

- sources of information required for completion

- decision point within the criminal justice system that the risk screening/assessment is aiming to inform

- sensitivity and specificity

- type and number of items.

3.3.1 Sources of information

Instruments vary according to the sources of information required for completion this can include police records, court records, school records, ratings by informants such as parents, teachers, social workers, treatment providers or diagnosis by mental health professionals. The type of information required can impact on an instruments utility in different contexts and/or for different users.

Some instruments are based on ratings by experts or clinicians and are therefore only suitable for those with appropriate professional qualification (e.g. PCL:YV is designed for registered psychologists). Other instruments are scored by youth justice practitioners following ratings provided by informants such as parents or teachers (e.g. the APSD – Antisocial Process Screening Device – is a 20-item rating scale of children six to 13 years completed by one teacher and one parent). There are a few instruments where ratings are based on self-report provided by young offender (e.g. self-report version of APSD or Psychopathy Content Scale [PCS] - 20-items; or the Young Person [YPI] Psychopathic Traits Inventory [50-items]). More often instruments are based on information from several sources, file records, youth self-report, expert opinion, parent and school.

The availability of specific types of information will determine the utility of particular instruments. For example, registered psychologists will not always be on hand to administer instruments such as the PCL:YV. Also, data from certain sources are not always readily available. This was an issue for a tool designed to screen for risk in incarcerated young offenders in New Zealand, the Risk of Serious Youth Offending tool (RSYO). Completion of the RSYO required data to be collected from a parent or other designated adult, however, this was difficult to obtain due to difficulties in locating designated individuals and/or reluctance of these informants to participate (Wilson and Rolleston, 2004).

3.3.2 Decision point

The availability of information, and the interests and experience of those conducting the assessments is influenced to a large extent by the point at which the screening or assessment is to be completed. Vincent et al. (2009) note that risk screening/assessment instruments tend to be tied to particular decision points, with different points being associated with different

questions and assessment needs. Examples of different points and different information requirements include:

- **At first intake or contact with police** – is youth appropriate for diversion?
- **Prior to court if charged** – does youth need secure pre-trial detention?
- **At court –** what is appropriate placement, security level, treatment plans?
- **If incarcerated –** what is the risk of suicide and/or violence? What type of interventions are most appropriate?
- **On community re-entry –** what is appropriate monitoring and intervention?

The different questions and assessment needs at each point will vary according to a jurisdictions' particular youth justice system. Compared to other jurisdictions New Zealand Youth Aid Officers have a high degree of autonomy in deciding what level of action will be taken for any young offender they come into contact with (McLaren, 2009). For example data published by the Ministry of Justice suggests that in over a third of cases police decide on Alternative Action.[8] This requires police officer's to develop an Alternative Action Plan, and hence information on level of risk and criminogenic needs is relevant.

In contrast, the United States has a greater focus on pre-trial placement decisions (Vincent et al., 2009) such as whether the youth should be placed in detention, electronic home monitoring or not detained at all. These decisions are influenced by information on a young person's risk of recidivism and failure to appear in court. In this situation ability to predict recidivism in the short-term is more important than information on treatment needs (Kalb, 2006).

Vincent and colleagues also point out that there tend to be different populations of young offenders at each decision point and so different risk factors may be relevant. The group of young offenders initially dealt with by police will be diverse and include a large number of low risk first time offenders and a smaller group of serious and persistent offenders. The total volume of young offenders being dealt with decreases at each subsequent stage of the criminal justice system. For example in New Zealand, of all youth (14 to 16 years) apprehended by police in 2007, just under a third (28%) were prosecuted.[9] Of all the proved cases in Youth Court just 13.4% received a supervision with residence order, while 23% of those convicted in the District or High Court were imprisoned (Duncan, 2009). While the numbers at each stage are decreasing, the proportion of youth considered to be at high risk of recidivism increases.

A tool needs to have the right items to be able differentiate between the particular type of young offenders of interest (e.g. low, medium, high risk) at the point where the tool is being applied. Kalb (2006) cites research that suggests prediction of recidivism becomes increasingly difficult as the base rate differs from 50% (i.e. the actual rate of reconviction of the youth being screened is differs greatly from 50%).

In general, Weatherburn et al., (2007) suggest the earlier in the youth justice process that screening is undertaken the better, particularly if early intervention is the main goal. However, as noted above the ability to screen requires that information about the young person is available.

---

[8] Alternative Action given to 37.1% of 14-16 year olds and 44.9% of 10-13 year olds (see Duncan, 2009)

[9] Of those prosecuted, 26.5% were proved in youth court, and 3.9% convicted in District or High Court.

Weatherburn et al. (2007) were interested in screening young offenders based on file information, which meant that the earliest point they could carry out a screen was after a supervised community-based order had been given in court. Prior to this point information recorded in files was insufficient.

Based on research in the United States, Smith and Aloisi (1999) suggested screening young offenders at their second arrest was most efficient. This was based on the knowledge that most first time young offenders did not go on to re-offend so screening them all was not an efficient use of resources.

The decision point will also determine who is likely to be administering the screening/assessment. Screening/assessment at first contact is more likely to be carried out by police officers, but depending on how the young offenders is processed, subsequent screening/assessments are more likely to be carried out by youth justice social workers, treatment providers and/or correctional officers. A screening/assessment tool must therefore be applicable to the skill levels, areas of expertise and time available of the individual conducting the screening/assessment.

3.3.3 Number and type of factors

The content of a tool will again depend on its purpose (e.g. to identify a personality trait, risk of recidivism, providing case management information and/or measuring change), however, Schwalbe (2007) suggested the most common risk factors found in risk screening/assessment tools are:

- offending history
- substance abuse
- family problems
- peer delinquency
- school related problems.

In terms of the length of the instrument or number of factors, the Youth Justice Board (2008) noted that if statistical prediction of recidivism is all that is required – a short actuarial tool is perfectly adequate.

In Australia, Weatherburn et al. (2007) found that they were able to determine which juveniles will re-offend with a fair degree of accuracy using just four pieces of information that are routinely collected by the government. These were a mix of static and dynamic factors:

- age at time of re-offending
- not being at school
- having been suspended or expelled from school
- having several prior contacts with the criminal justice system.

Interestingly research in the United States carried out by Krysik and LeCroy (2002) found a five-item actuarial model that included both dynamic and static factors performed better in predicting recidivism than a than nine-item scale based on only static factors, although, the authors cautioned assessment error may have undermined the accuracy of the nine-item formula. Similarly, research on an early version the UK's ASSET found a shorter version was more accurate in predicting recidivism that the longer version. This was tested by removing items that

did not significantly predict recidivism. Interestingly the final version of the instrument included all items. This was because some items were judged to have clinical relevance even if they did not contribute to predicting recidivism (Baker et al., 2002).

Here in New Zealand, the Department of Corrections uses an actuarial model to predict probability of recidivism and re-incarceration for adult offenders that achieves a high level of accuracy based on the following eight static variables:

- current age
- gender
- age at first conviction
- number and seriousness of convictions
- number and type of previous sentences
- amount of time spent in custodial sentences.

If, however, the aim is also to identify areas for intervention and to measure change over time, then the tool will need to incorporate a wider range of variables (Youth Justice Board, 2008). Tools commonly used for this type of case management include the YLS/CMI which consists of 42-items that assess both static and dynamic factors, while in the UK the ASSET is used which has 106 items.

The Youth Justice Board (2008) suggests that because adolescence is a period of substantial development and transition, frequent and repeated assessments are warranted. If re-assessment is to be carried out than the inclusion of dynamic factors is essential. An instrument based on static factors alone will produce the same risk score regardless of any change in a young persons personal circumstances or if criminogenic needs have been addressed through intervention.

The measurement of particular constructs also appears to benefit from the use of more detailed instruments. Murrie and Cornell (2002) compared three briefer screening tools for psychopathy with a full assessment using the PCL-YV which takes clinicians 60-90 minutes to complete. The found substantial limitations with the performance of the briefer screening tools.[10]

3.3.4   Sensitivity and specificity

When tools are used to classify youth into high and low risk offenders an important consideration is the relative sensitivity and specificity of the tool. The sensitivity and specificity of a tool relates to its ability to minimise two types of error. A highly sensitive tool minimises the number of 'misses' or 'false negatives' (i.e. the number of young offenders that the tool identified as being at low risk of recidivism but turn out to be a recidivists). A tool with good specificity, minimises 'false alarms' or 'false positives' (i.e. a young offender is incorrectly identified by the tool as recidivist).

---

[10]   The briefer screening tools were reasonably accurate in predicting PCL-YV high scores (67-82% correctly classified) but had a relatively low sensitivity for screening purposes (i.e. ability to detect true positives), with at least a few high-psychopath youth receiving low screening scores who therefore, would be missed (Murrie & Cornel 2002).

There are costs and consequences associated with both types of errors and tools are often designed to give priority to one or the other by lowering or highering the cut-off threshold. Weatherburn et al. (2007) explains that lowering the threshold for classifying a young person as a likely recidivist will increase the number of 'hits', but will also increase the number of 'false alarms'. The cost of a 'false alarm' is the money wasted when a juvenile offender, who would have ceased offending anyway, is subjected to an unnecessary intervention. If interventions are expensive and places limited, this type of cost may be unacceptable, in which case the threshold would be raised to decrease the number of false alarms but increase the number of 'misses'. The cost to consider in this situation include the impact on victims of subsequent offending and the costs associated with processing the new offending through the criminal justice system that may have been avoided if the young person had been correctly identified and given an effective intervention.

In general, if screening is followed by further assessment, priority should be given to a tool that is highly sensitivity (Murrie & Cornell, 2002). This will minimise the misses or false negatives and ensuring all those youth who are recidivist are identified. Higher numbers of false alarms or false positives are less problematic as subsequent more detailed assessment should then identify and divert these young people (i.e. the low risk offenders that will not go onto re-offend and do not require intensive intervention). However, if screening is not followed by a further assessment high rates of false positives become more of a concern due of the costs of large numbers of youth being inappropriately referred to a resource intensive intervention.

3.3.5 Other attributes

Vincent et al. (2009) emphasise the importance of tools being empirically based. They list the following as the minimum requirement.

- **Standardisation:** having a manual, scoring rules and detailed item descriptions.

- **Reliability:** evidence that the tool can produce consistent results, across time and raters.

- **Validity:** measuring what the tool purports to measure (e.g. recidivism and/or criminogenic needs).

- **Empirically based risk factors: containing factors demonstrated to have a statistical** association with future criminal behaviour. They also suggest if the tool is measuring particular constructs (e.g. psychopathy, impulsivity) then it should have a known internal structure and, if score-based, should have normative data against which to assess the relative standing of the client.

More details on how to assess the validity and reliability of an instrument will be included in the next section on evaluating risk screening/assessment tools.

_____

3.4   Limitations of risk screening/assessment instruments

While there are many advantages to standardised risk assessment it is also important to be aware of some of the inherent limitations of this type of assessment.

It is argued that this type of assessment is over simplified and focuses too much on individual factors while not taking sufficient account of complex social problems (O'Mahony, 2009; Webster et al 2006). According to O'Mahony's (2009) the risk factors prevention paradigm, upon which risk screening/assessment instruments are based, in many ways acts as an obstacle to a fuller understanding of, and more effective response to, youth crime. He argues that this results in an over-simplification of the key issues, failing to account properly for key facets of youth justice such as personal agency, socio-cultural context, psychological motivation and human rights.

Others also argue that current risk screening/assessment instruments miss a big part of the context in which the offence occurs; such as whether the offending occurs with associates or alone or whether the youth were under the influence of substances (Thompson, 2003). In addition, Thompson (2003) notes that developmental issues are not well addressed in most risk screening/assessment instruments.

Reviews of practitioners' perspectives by the Youth Justice Board (2008) suggest many practitioners regard standardised assessment as diminishing practitioners' ability to exercise skilled and professional judgement. Comments from youth justice practitioners suggested they perceived risk screening/assessment instruments to be 'de-skilling', there were also concerns expressed over the time structured assessments took to complete (Youth Justice Board, 2008).

A final and important limitation of risk screening/assessment instruments is that they categorise young offenders on the likelihood (or probability) of them being reconvicted within a certain timeframe. At issue is the divergence between probability certainty that a particular young person will re-offend (Shader, 2003). Based on the principles of probability, in a sample of 100 youth that have been classified as a group as having an 80% chance of re-offending (i.e. high risk) – 20 of these youth will not re-offend (assuming the tool is 100% accurate). The Youth Justice Board (2008) cautions tools cannot be accurate in all cases and there will be times when professional judgement will need to override an instrument's score (Youth Justice Board, 2008; Vincent et al., 2009).

3.5   Summary

This chapter has presented the main methods that are used to predict risk of recidivism in youth and introduces the different types and attributes of risk screening/assessment instruments that have evolved. Understanding the attributes of different type of instruments will be useful in the evaluation of those of the YORST.

More recently there has been growing recognition of the importance of using screening/assessment instruments that include empirically-based risk factors that assess both risk and need. These are referred to as third generation risk assessment instruments and include both static factors and dynamic factors.

The number and type of factors included in an instrument will depend on its purpose. If statistical prediction of recidivism is all that is required a short actuarial tool, with predominantly static variables, is likely to be perfectly adequate. However, if the aim is to identify areas for intervention or to measure change over time, than it will require a more detailed assessment instrument with priority given to dynamic factors.

Instruments vary according to the sources of information required for completion. This can include police records, court records, school records, ratings by informants such as parents, teachers, social workers, treatment providers or diagnosis by mental health professionals. An instrument must be feasible to complete with available resources, skill levels and type of experience of those carrying out the assessment.

The instrument must also be applicable to the decision point it is aiming to inform (e.g. when young offender first comes into contact with police, prior to or following court if young person is charged, at beginning of an intervention or custodial sentence or intervention, and on community re-entry). Information requirements will vary at different points and may be jurisdiction specific. For example, compared to other jurisdictions, New Zealand Police Youth Aid Officers have a greater degree of autonomy in deciding the appropriate level of action police for any young offender they come into contact with. This means information on risk and criminogenic needs will be relevant.

When tools are being used to classify youth into high and low risk offenders another important consideration is the relative sensitivity versus specificity of the tool (its ability to minimise false negatives compared to false positives). This relates to a tools associated threshold for classifying a young person as high risk and therefore in need of intensive intervention. If the threshold is too low there will be a number of young offenders being diverted to treatment unnecessarily. If it is too high, then there are likely to be chronic serious recidivists that are not identified who will not receive the intervention they need to reduce the likelihood of further offending.

There is growing recognition over the advantages of standardised risk screening/assessment (e.g. increased accuracy, consistency, equity and objectivity), however, there are also some limitations associated with this type of assessment. Some argue that structured risk screening/assessment can result in over-simplification of complex social problems. Others point out that that they can diminish practitioners' ability to exercise skilled and professional judgement. A final concern, it the failure to properly understand that risk classification is based on probability, however, but being at high risk does not make offending a certainty.

4 Evaluation of risk screening/assessment tools

The previous chapter reviewed the range of different types of risk screening/assessment instruments that have evolved over time, together with their varying attributes. A convincing rationale was presented for implementing structured forms of screening and assessment in order to increase consistency and reduce potential bias in youth justice decision-making. However, if a tool is used inappropriately or it is invalid it can have a deleterious effect on the validity of the decision-making (Vincent et al., 2009). This chapter briefly reviews important considerations in selecting and/or developing such tools including whether the tool has adequate validity and reliability, if there adequate policies, protocols and training supporting the tool and if the tool is appropriate for the local context in which it will be used.

4.1 Assessing validity/reliability

For a risk screening/assessment instrument to be able to improve the quality of youth justice decision-making it is must be both valid and reliable.

- **Validity** refers to the degree to which an instrument accurately reflects or assesses what it is intending to measure.

- **Reliability** is the extent to which a test is consistently able to produce the same result across time and across raters.

4.1.1 Validity

There are three main types of validity to consider, these are:

i) **Content validity:** which relates to whether the tool has the right content and includes the right items to be able to measure what it is intending to measure). For example, the content of a risk screening/assessment tool designed to guide pre-trial placement decisions will need information on risk of harm to others and past histories of failure to appear, while a tool designed to inform appropriate intervention will require content to assess criminogenic needs. If a tool is required to measure change in risk of recidivism, then dynamic factors rather than static factors are required. Content validity is typically assessed by a panel of experts evaluating test items against test specifications.

   Content validity is different from **face validity** which rather than being rigorously tested, requires more of an intuitive judgement by those using the instrument that it 'looks valid'. Face validity is important to ensure there is maximum uptake of a tool by practitioners (i.e. practitioners can see the value of the tool and chose to administer it).

ii) **Construct validity:** is another form of validity important for psychometric instruments that include a number of items that measure a number of sub-scales or particular constructs (e.g. psychopathy or substance abuse). In this case the items used to measure each construct or sub-scale should be in agreement with those measures they are theoretically related to (convergent validity), but not highly correlated were those designed to measure theoretically different concepts (divergent validity). This is type of validity testing is

applicable to many of the more detailed risk screening/assessment instruments such as the YLS/CMI, PCL:YV, ASSET, but is not applicable for screening tools like the YORST that have very few or just one item relating to a particular construct (e.g. family relationships, drug and alcohol use).

iii) **Criterion validity (concurrent or predictive):** is the type of validity most frequently tested and referred to in relation to risk screening/assessment instruments for young offenders. This measures the degree of association between an instrument and what it is intending to measure (e.g. risk of recidivism).

- **Concurrent validity** is established by demonstrating agreement between the tool in question and another risk screening/assessment tool with established validity (e.g. YLS/CMI). The two tests are completed at the same time and the level of agreement between the validated test and the tool being developed assessed.[11]

- **Predictive validity** is perhaps the most commonly discussed type of validity in reference to risk screening/assessment tools and relates to their ability to predict a criterion variable (i.e. recidivism). A tool is typically developed retrospectively on a 'construction' sample. However, once developed, it is critical for predictive validity to be cross-validated, and tested prospectively on a different 'validation' sample of young offenders to that from which the tool was first developed (Kalb, 2006; Gottredson & Snyder 2005; Schwalbe, 2007; Vincent et al., 2009).

Vincent et al. (2009) reviewed a number of statistical tests to assess predicative validity and recommend Cox Proportional Hazards Regression, Survival Analysis and the most commonly cited Receiver Operating Characteristic (ROC). The area under the ROC curve (AUC) is an index of a tool's overall accuracy and produces scores ranging from 0 to 1 (0.5 indicates a chance-level accuracy; below 0.5 indicates below-chance accuracy with tools incorrectly classifying the majority of offenders; above 0.5 indicates accuracy above-chance). AUCs for an acceptable screening tool would be between 0.70 and 0.90 (Swets, 1988: cited in Vincent et al. (2009).

It is not uncommon for newly developed tests to be subjected to both types of criterion validity testing. Testing concurrent validity is time and resource efficient. It also eliminates the problem of confounding variables such as impact of the number and type of interventions a young person has been exposed to on subsequent reconviction. However, this method assumes that the tool with established validity is appropriate for the local testing setting. Instruments with established validity that are commonly used to test concurrent validity include the YLS/CMI, PCL:YV, SARVY (see Vincent et al., 2009).

In contrast, testing predictive validity does provide a context specific measure of a tools accuracy. Predictive validity also allows a tool to be tested against a measure most relevant for the tool (e.g. re-arrest, re-conviction or reincarceration). The main draw backs are the time taken to carry out such testing as it requires a sufficient follow-up period after implementing the tool to

---

[11] Concurrent validity is sometimes referred to as convergent validity (see construct validity).

collect recidivism data (minimum period would be six months, but preferably it would be 12 or 24 months). The other problem is eliminating the effects of confounding variables that occur during this follow-up period (e.g. type and number of interventions a young person has received).[12] If tools are administered post intervention (e.g. prior to release from custody, or at the end of a court order or treatment programme), then the confounding impact of interventions can be minimised. In this way the risk prediction and follow-up period where reconviction is assessed occurs after the intervention. However, unfortunately most instruments are administered prior to sentencing or any intervention in order to inform case management.

In assessing the predictive accuracy of an instrument consideration should also be given to the validity of the outcome measure. For example, when using official reconviction statistics, the instrument is actually being tested against who gets caught (and/or arrested or charged) and not necessarily who re-offends. Kalb (2006) also points out that when assessing a tools ability to predict seriousness of offending, it is important to note that serious offenders also engage in high levels of less serious offending. This means in assessing an instruments ability to predict seriousness of offending, it will be necessary to take into account that based on probability the individual will be convicted a less serious crime next.

Rather than selecting one type of validity testing, an instrument ideally is subjected to a range of tests that take into consideration a number of different validity criteria. For instance, Baker et al., tested the ASSET against the following range of validity criteria (Baker et al., 2002; 2005):

- accuracy in predicting whether someone is reconvicted or not (percentage predicted correctly and Area Under the Curve (AUC) analysis)
- accuracy across the instruments range of scores (ASSET scores divided into five equally sized bands from low scores through to high scores and rates of reconviction compared across bands)
- accuracy in predicting frequency and seriousness of reconviction
- prediction accuracy for selected population sub-groups (ethnic minorities, females and younger offenders).

Valid measure of change: Instruments that include dynamic factors if re-administered overtime theoretically have the ability to detect changes in levels of risk of a young person. In a review of risk/needs tools for antisocial behaviour and violence among youth, Vincent et al. (2009) observed that few tools contain dynamic factors that would lend themselves to reassessment and those tools that are designed to measure change lack evidence with respect to their ability to do this. This was viewed as a serious issue, as labelling a youth as high risk without providing a method for measuring any reductions in risk that may have occurred over time could have a deleterious effect.

The ASSET is one of the few instruments that has carried out an evaluation of its ability to detect changes over time. Preliminary evidence, has found the tool to be sensitive to changes over time (both negative and positive), and decreases in scores of those youth on community

---

[12] Kalb (2006) proposes that another and perhaps more appropriate use of risk screening/assessment instruments is rather than predicting who will be reconvicted, to predict for an individual offender which available treatment option will most likely result in a successful intervention. Analysis would involve comparing risk assessment scores to level of responsiveness to the varying types of interventions.

sentences (but not custodial sentences) were found to be associated with reduced likelihood of reconviction (Baker et al., 2005).

4.1.2 Reliability

It is also critical that a test is consistently able to produce the same result across time and across raters. The two primary forms of reliability typically referred to are:

- **Stability reliability (or test-retest reliability):** which is the agreement of scores produced by an instrument over time. For example, if a tool was administered twice on the same young offender on consecutive days would the same results be achieved? NB: It is important that the re-test occur within a time period where dynamic items on the instrument would not have been expected to change.

- **Inter-rater reliability:** which is the degree of agreement among two or more individuals using the same instrument. If there is not inter-rater agreement either the tool is defective or the raters' need to be re-trained in its use. Again, there are a number of different statistical tests that can be used to evaluate inter-rater reliability; Vincent et al. (2009) recommends intraclass correlation coeifficients (ICCs) which should be 0.70 or higher.

Vincent et al. (2009) suggest if a risk tool is based on self-report, then the focus should be on test-retest reliability, however, if a tool relies on subjective ratings by practitioners then inter-rater reliability is more important. The later case is more common with risk screening/assessment instruments.

There are a number of ways that tests of inter-rater reliability can be approached. The most common is a matched pairs approach. This involves independent assessments, by two practitioners, of the same young offender. This may require the two raters to attend together a series of interviews with young offenders and examine the same file information before completing the instrument. It can also involve the young person going through two sets of interviews. Inter-rater reliability is then determined by comparing the level of agreement between these two raters. This matched-pairs approach was the method used in initially assessing the inter-rater reliability of the YLS/CMI (Hoge & Andrews, 2002). Reliability can be established across different professional groups such as mental health professionals and probation officers (see Schmidt, Hoge & Robertson, 2005), or between researchers trained as raters (see Catchpole and Gretton, 2003).

The inter-rater reliability of UK's ASSET tool was assessed using an alternative approach. Researchers developed four young offender case studies that varied according to gender, ethnicity and age, and reflected different offending scenarios. Videos were produced of drama students role playing being interviewed by youth justice practitioners, and this was supplemented by fictional file information that would be required to complete an ASSET (see Baker et al., 2005). The case studies were then sent out to a number of youth justice practitioners for their assessment and inter-rater agreement was compared. The analysis assessed inter-rater reliability across raters and different offender characteristics. Some discrepancies were revealed based on gender, with raters of the female offender case study appearing to allocate ratings based on welfare needs rather than the risk of re-offending. The researchers reported that the exercise had beneficial training outcomes, however, it appeared the exercise was quite onerous for youth justice practitioners, participation was voluntary and

response rates by youth justice practitioners returning their assessments was low. A further limitation of this approach was that the fictional case studies did not allow the instrument to be tested in a real life situation where relevant information would not presented in such a standardised and efficient manner.

Previous to this study inter-rater variations using the ASSET was tested by comparing total ASSET scores to a static score for each young person based on their offending history (Baker et al., 2002). This allowed consistency to be compared on a large number of raters, across different professional groups, different stages of the youth justice system, different levels of experience and across different regions. A limitation of this approach is that it assumes that the static score correlates with the ASSET rating, which is not necessarily the case. In come cases a high ASSET score could be achieved by a young offender with no previous offences.

Implications of low inter-rater reliability

If an instrument is found to have low inter-rater reliability this can mean one of two things: (i) that there is a problem with the instrument; or (ii) there is a problem with the way raters have been trained in the use of the tool or accompanying instructions.

It could be the tool has items that are too subjective, that wording need clarification or that there is a problem with the scoring system. Baker et al. (2005) found it was possible to improve the inter-rater reliability of the ASSET by converting the five-point scoring system into a yes/no option.

Alternatively, improved training and/or instructions may be required (Schwalbe, 2007). Analysis of the reliability of the ASSET found some raters found it difficult to distinguish between the problems in a neighbourhood and the impact of these on the offending behaviour of a specific young person. The researchers suggested this could be addressed through training (Baker et al., 2005).

In the field of family violence, it has been found that inter-rater reliability among those screening for risk of family violence can be increased through raters being brought together to discuss why they would give particular ratings. This approach has been referred to as 'calibration' of raters (Trochim, 2006).

4.2   Other evaluation considerations

4.2.1   Evaluation of policies, protocols and training

The role that appropriate training and supporting documentation play in assisting a tool to be administered reliably suggest that evaluation of these aspects be carried out. It is important to know the tool is being implemented as intended, that it is user friendly and easy to use. Regardless of its validity, if is incorrectly implemented will not achieve valid or useful results.

This could involve a review of documentation by researchers, together with consultation with practitioners on their experience, and in particular any problems, in using the instrument. Field observations can also be a useful way to assess implantation issues. This approach to evaluation is supported by Schwalbe (2007) who recommends that youth justice agencies should periodically evaluate the way risk screening/assessment instruments are implemented.

Another important aspect for consideration is the level of uptake of an instrument by practitioners. Again a tool may have perfect validity and reliability, but if it is not valued by those practitioners responsible for implementing it. The Youth Justice Board (2008) in reviewing practitioners perspectives on assessment tools, noted the context in which tools are used and the surrounding organisational culture can impact on the value practitioners see in such tools. Another critical factor in the degree of uptake of the tool is the degree to which it is user friendly (e.g. appropriate length, easy to access information to complete questions). If it is difficult to complete, uptake will be low and its potential value may be seriously compromised. This is particularly true if the aim of the instrument is to collate data on a regional or national level.

Baker et al. (2005) also noted that for a tool to be seen to be useful, it is helpful if it aligns with related processes. Evaluation of ASSET found practitioners were not incorporating ASSET findings into intervention plans (Baker et al., 2005). In response, the researchers suggested closer integration of the ASSET and the intervention plan may help. Hence, consideration of how a screening/assessment tool aligns to associated youth justice processes should also be part of a an evaluation.

4.2.2   Appropriate for the context

A final consideration is whether an instrument is appropriate and valid for the local context. A recent study by Miller and Lin (2009) suggested generic or 'off the shelf' tools are less accurate in predicting recidivism than a locally developed risk screening/assessment tool. Although, some 'off the shelf' tools such as the YLS/CMI have been demonstrated to be valid across jurisdictions (see Olver et al., 2009). As noted earlier, of particular relevance in New Zealand is the importance of tools to be applicable for Māori young offenders.

As reviewed in the previous chapter to be appropriate for the context, the tool must also provide information that is relevant to the specific assessment question, which relates largely to the setting and decision point. It must also be feasible to complete the tool given the resources available and the skill set of their staff (Vincent et al, 2009).

Instruments vary in the level of expertise of the examiner and the amount of information required to complete the assessment. Qualified psychologists or social workers will be trained to complete comprehensive assessments using a range of information (Vincent et al., 2009). Other non-clinical youth justice practitioners such as police officers or correctional facility staff may have limited time to conduct assessment and may not have received appropriate training to assess mental health status.

Hence, evaluation of a tool should include whether it is it must be feasible to complete with available resources and level and types of skills, and whether it is relevant the specific youth justice decision point (e.g. arrest, court, intervention) and local jurisdiction.

4.3   Summary

This chapter has reviewed important aspects associated with evaluating the appropriateness of risk screening/assessment tools, and what key considerations might inform subsequent research to validate the YORST.

For an instrument to be to be able to improve the quality of youth justice decisions it must be both valid and reliable. Ideally a newly developed instrument will be subjected to a range of validity and reliability tests.

To be valid, an instrument must be shown to have appropriate content to be able to measure what it is intending to be measure, if there are sub-scales, that there is evidence that they measure distinct constructs, but perhaps most importantly that the instrument has proven criterion validity (i.e. that is accurately measures what it is intending to measure). There are two types of criterion validity, concurrent (demonstrated to agreement with another established instrument) and predictive (tested against an instrument with established validity) and predictive (tested against the variable it is intending to measure, i.e. recidivism). Instruments with established validity that are commonly used to test concurrent validity of newly developed instruments include the YLS/CMI, PCL:YV, and SARVY. Predictive validity, should be tested prospectively on a 'validation' sample of young offenders, different to the 'construction sample' from which the tool was first developed. Other concerns in testing predictive validity are selecting an appropriate outcome variable and how to account confounding variables such as the impact of interventions on recidivism, if they have occurred following the risk screen/assessment.

Equally important as validity, is the reliability of the instrument, the extent to which it is consistently able to produce the same result across time and across raters. Most risk screening/assessment instruments are based on subjective ratings by practitioners so inter-rater reliability is most important. This is most commonly tested using a matched pairs approach, where independent assessments are made by two practitioners of the same offender and compared for their level of agreement.

In addition to establishing the validity and reliability of the instrument it is also important to evaluate the policies, protocols and training that surround the tool and to carefully consider if the tool is appropriate for the particular context in which it is to be used.

5 Risk screening/assessment tools

This chapter begins with a review of youth risk screening/assessment tools that have been used in New Zealand. This is followed by findings from studies that have reviewed the performance of risk screening/assessment instruments that are used in other jurisdictions. Specific details of some of the more widely used instruments, together with details of their validity and reliability are presented in Tables 4.1 to 4.2.

5.1 Risk screening in New Zealand

In the past, there has been limited risk screening of young offenders in New Zealand. This prompted the development of the YORST and its predecessor the ARNI. However, other risk screening/assessment tools do exist and we present below information on other risk screening/assessment tools that have been developed and/or trialled in New Zealand.

5.1.1 Adult Offenders

ROC*ROI: As referred to earlier, the Department of Corrections has successfully integrated an actuarial model called ROC*ROI into their adult offender management system. The model was first developed in 1994 and aimed to predict risk of reconviction (ROC) and risk of re-incarceration (ROI) based on a range of static factors from an offender's criminal history (Bakker et al., 1998). A computer generated score assesses the probability (0.0 to 1.0) of a particular offender incurring new convictions (ROC) and being imprisoned (ROI). Factors used in the calculation of individual risk scores include:

- current age
- gender
- age at first conviction
- number and seriousness of convictions
- number and type of previous sentences
- amount of time spent in custody.

AUC analysis carried out on the original model found good predictive validity AUC=.76 (Bakker et al., 1999: cited in Wilson & **Rolleston**, 2004). We understand this model has subsequently undergone some minor revisions including the removal of ethnicity as a risk factor.

Unfortunately criminal history data does not include youth court data and, therefore, the model has limited utility for young offenders who have not entered the adult criminal justice system, and even less for the majority of young offenders who come into contact with police and are dealt with through alternative action. Wilson & Rolleston (2004) tested its predictive ability on a sample of young offenders aged 16 to 19 years who were serving time in Department of Corrections' Youth Offender Units and found it to have a moderate significant correlation with previous total convictions (r=.68). This was higher than other more established youth risk assessments including the PCL-YV (r=.46) and the YLS/CMI (r=.44).

5.1.2 Risk of recidivism in young offenders

In addition to the YORST and its early version the ARNI, our review only identified two risk screening/assessment **instruments**, and one actuarial model to predict risk of recidivism designed specifically for New Zealand young offenders.

- **Risk of Serious Youth Offending (RSYO)**: The Department of Corrections also carried out development work on an screening/assessment tool called the Risk of Serious Youth Offending (RSYO). It was designed to estimate those who would become chronic life-course persistent offenders (LPC) compared to adolescent limited (AL). The tool took one to two hours to complete and required an interview with the young offender and a nominated significant other for cross-checking (Wilson & Rolleston, 2004). Wilson and Rolleston's (2004) evaluation of ROC*ROI referred to above, also evaluated the validity of the RSYO. They found significant moderate correlations with the PCL:YV and YSR/CMI (r=.63, r=.70),[13] but in general found it performed less well than other measures predicting reconviction tested. It was also noted that interviewing significant others was difficult to achieve as they often proved hard to locate or reluctant to participate. It appears development work on this tool did not continue.

- **Youth Risk Screen (YRS):** The YRS was developed as part of an initial screening process for a New Zealand young offenders' programme called Reducing Youth Offending Pilot (RYOP).[14] Details of the tool are included in the evaluation report of the programme (Grace et al., 2006). The researchers described that the YRS is administered as a form of eligibility screening (to ensure that the young person meets programme high risk eligibility criteria) prior to acceptance on the programme. The YRS is designed to capture the following information:

    □ prior offending
    □ early evidence of Conduct Disorder/Oppositional Defiant Disorder
    □ Attention-Deficit/ Hyperactivity Disorder.

According to Grace et al. (2006) the YRS was based on Moffit's 1993 research on chronic offenders. The authors reported **that** the YRS was locally developed and had not been validated prior to the evaluation of the RYOP programme. Grace and colleagues tested a modified version of the YRS that could be completed using records held in CYF's CYRAS database and found it was significantly correlated with three measures of recidivism (r=.25 to r=.34) and was found to have moderately good predictive accuracy for both males and females.

---

[13] The correlation coefficient 'r' is a measure of the strength of the association between two variables. Values range from –1 to +1, indicating perfect negative correlation at –1, absence of correlation at zero, and perfect positive correlation at +1.

[14] The RYOP programme is based on multi-systemic therapy (MST) an intensive intervention that has been well-established internationally as being effective. It was trialled between 2003 and 2006 and then evaluated.

- **Actuarial models:** As part of the evaluation of the RYOP programme, three actuarial models were developed to predict recidivism in a sample of 500 young offenders (Galletly, 2006; Grace et al., 2006).[15] All offenders included in the research had received a CYF youth justice intake in 2002. The factors included in the most successful model (M2) have been reviewed in chapter three. Below is a summary of the accuracy of all three models on the sample used in their development and also for M2 and M3 on a new validation sample.

  - M1 predicted CYF youth justice intakes: development sample AUC= .713

  - M2 predicted prosecution of any re-offence: development sample AUC=.73; cross validation sample AUC= .738

  - M3 predicted prosecution of any serious re-offence: development sample AUC=.723; cross validation sample AUC= .678

- The results of M2 are similar in accuracy to that achieved by widely used and validated instruments such as YLS/CMI, PCL-YV, SAVRY and ASSET.

5.1.3 Other screening of New Zealand young offenders

In 2000 CYF introduced a system of screening and assessment tools for social workers to administer on children aged 12-13 years, and young people between 14 and 16 years (CYF, 2006). The tools aimed at providing a more **systematic** process of screening and assessment to improve youth justice decision-making. The tools are used across Care and Protection, Youth Justice and Residential Services. The package includes the following tools:

- **Cage:** a screening tool for drug and alcohol use (asks about recent alcohol and drug use followed by a further four standard questions).

- **Kessler:** is a screening tool for psychological distress (6-items).

- **Suicide tool:** screens for risk of suicide (primarily clinical judgement but provides three prompts for social workers to ask young person).

- **The Suicide Risk Assessment and Suicide Risk Management Plan:** intended to support the Social Worker to investigate the level of suicide risk and describe the steps they will take to respond to that risk.

- **Wellbeing assessment:** is used to assess the needs and strengths of a young person and their family. The tool is primarily designed for youth justice, but can and is used in Care and Protection. It covers the following domains:

  - pattern of offending
  - family/whānau environment
  - education/employment
  - physical wellbeing

---

[15] A model that could accurately predict recidivism could be used to evaluate if the programme reduced re-offending below what was predicted by models. A similar strategy is used to evaluate the effectiveness of Department of Corrections' rehabilitation programmes.

- ▫ emotional well-being
- ▫ attitudes
- ▫ social interactions and peer relationships
- ▫ spiritual and cultural identity.

We were unable to **locate** more details about its development or validity or reliability. The Werry Centre (2009) reported that the there is little evidence on reliability or validity of the Cage-Kessler with adolescent populations.

Current developments: CYF are currently in the process of reviewing the package of tools used to screen and assess young people. Work is underway to update the package of screening tools Social Workers use to help identify substance abuse, mental health concerns and suicide risk. Specific consideration is being given to **introduce** screening tools that have proven validity and reliability and are recognised by the Health sector. The review of the Wellbeing Assessment is complete and it is due to be replaced later this year. The review found that the Wellbeing Assessment required updating to support CYF in their work with young offenders and meet the requirements of a fourth generation risk/need assessment. The new tool is currently under development. Key features of the tool include a summary of the young person's overall strengths, risks and needs, an indication of the risk of re-offending of future offending and an analysis of their offending behaviour (Personal communication with Child Youth and Family, 8/4/2010).

5.2 Risk screening/assessment tools used overseas

Standardised risk and need assessment and/or screening of offenders is well recognised as best practice (Andrews & Bonta, 1998; The Werry Centre, 2009; Vincent et al., 2009) and has resulted in the development of a number of standardised instruments. Summarises of key characteristics of some of the more widely used risk screening and assessment tools, together with available data on their validity and reliability are presented below in tables 5.1 and 5.2.

Before presenting details of the instruments that are designed to predict recidivism in young offenders, it is useful to consider the range of other tools that are available for youth justice population (those listed below have not been included in the tables that follow):

- Tools that are designed to predict early onset of offending prior to any conviction (i.e. identify children and young people at risk of re-offending and/or serious anti-social behaviour. This includes tools such as the Youth Justice Board's 'Onset' tools, or the Canadian Early Assessment Risk List for Boys (EARL-20B) and Early Assessment List for Girls (EARL-20G)

- Mental health screening and assessment tools designed for use with young offenders (for a review of these see Werry Centre, 2009)

- Tools designed to predict specific types of offending rather than general offending (e.g. Juveniles Sex Offender Assessment Protocol-II (J-SOAP-II) or Estimate of Risk of Adolescent Sexual Offence Recidivism Version 2 (ERASOR).[16]

---

[16] An exception to this is the SAVRY which was developed specifically to predict violent offending but has been shown to reliability predict general offending.

_____

- Actuarial models that rely on computer generated risk scores including Kalb's (2006) model to screen for risk of violence; Smith & Aloisis' (1998) model to predict second timers in the youth justice system; the UK's model to predict re-offending (Whiting & Cuppleditch, 2006); recidivism of juvenile offenders in New South Wales (Cain, 1996) and screening of juvenile offenders in New South Wales (Wetherburn et al., 2007).[17]

5.2.1   Tables of risk screening/assessment tools used overseas

Details of risk screening/assessment instruments primarily used to predict risk of recidivisms are presented in two tables:

- Table 5.1 include those with 20-items and/or take less than 30 minutes to complete, and require no clinical training to administer.

- Instruments in Table 5.2 are more comprehensive with more than 20-items and/or take greater than 30 minutes to complete. These may require clinical training to administer.

_____

[17]   Those that have been converted into an index that can be hand scored have been included.

Table 5.1: Brief/screening instruments

Details of instrument	Factors / content	Validity / reliability details
SECAPS Recidivism Risk Index – Putnins (2005) State/country of origin: South Australia Description: second generation, statistical model Factors: static and dynamic Length: 5-items Completed by: YJ practitioners Sources of info: file records plus youth self-report	Key purpose: estimating risk Risk classification: total risk score 0 to 28, calculated by assessors summing item components using Burgess method. Content: • prior proven offences • current age • age at first offence • alcohol and inhalant use frequencies • ADHD signs (feelings of restlessness, concentration, impulsiveness and boredom)	Predictive validity: • significant correlations with 6m post-release recidivism on validation sample of n=149, AUC=.71; r=.36 (Putnins, 2005) • risk prediction stable for different gender, age and ethnic groups (Putnins, 2005)
NCAR (North Carolina Assessment of Risk) - Schwalbe, Fraser, Day & Arnold (2004) State/country of origin: North Carolina, United States Description: second generation, based on OJJDP Model Risk Assessment Factors: static and dynamic Length: 9 items Completed by: YJ practitioners Sources of info: interviews, and file records	Key purpose: estimating risk Risk classification: cumulative risk score ranging from 0 to 30 (item responses summed). Content: • age at first offence • number of prior referrals • most serious prior adjudication • number of prior assaults • history of runaway from home or placements • severity of known alcohol or illegal drug use in previous 12m • delinquent peer associations • parental supervision	Predictive validity: • North Carolina adjudicated youth (n=14,719) recidivism measured by subsequent adjudication for an offence within12m f-up; AUC=.599 (Schwalbe, 2009) • predictive validity varied by gender, age and ethnic groups (Schwalbe et al., 2006) Reliability: • internal consistency: modest cronbach's alpha=.65 (Schwalbe et al., 2006)

Details of instrument	Factors / content	Validity / reliability details

ARNA (Arizona Risk/Needs Assessment) – Krysik & LeCroy, 2002
State/country of origin: Arizona, United States
 	
Description: second generation, statistical model
Factors: static and dynamic
Length: 5-items for first or second time offenders, 6-items for third time offenders
Completed by: YJ practitioners
Sources of info: file records
	Key purpose: estimating risk
	Risk classification: low, medium and high risk, scores ranging from 0 to 10. Computerised predicted probability and also cumulative risk scores using Burgess method.
	Different formula are used dependant on whether young person is a first-, second- or third-time offender.
	
	Content:
	• status offence - a
	• family relationship (current) - abc
	• assaultive behaviour (ever) - a
	• drug use in the past year - abc
	• current school enrolment - b
	• truancy or excessive absenteeism from school (ever) - abc
	• current behaviour and mental health issues - b
	• peer delinquency - c
	• runaway- c
	• prior complaints - c
	   a. Item included in first time offender index
	   b. Item included in second time offender index
	   c. Item included in third time offender index.
		Predictive validity:
		• youth referred to juvenile probation (n=29,711) recidivism measured by first delinquent complaint within 12m f-up; average across groups predicted probability AUC=.654, r=.25; cumulative risk AUC=.648, r=.24 (Schwalbe, 2009)
		• risk prediction stable for different gender, age and ethnic groups (Schwalbe, 2009)
		
		Reliability:
		• inter-rater reliability for 4-items requiring rating – 78.3% to 92.9% agreement (Krysik & LeCroy, 2002)
		• internal consistency: modest alpha=.54-.60 (Schwalbe, 2009)

Details of instrument	Factors / content	Validity / reliability details
APSD (Antisocial Process Screening Device rating scale) - Frick & Hare (2001) cited in Douglas et al (2008) State/country of origin: Canada Other versions: Previously Psychopathy Screening Device (PSD), also APSD –self-report tool same 20-items but re-worded in the first person for self-report Description: second generation, based on OJJDP Model Risk Assessment. Factors: personality Length: 29-items Completed by: teachers or parents Sources of info: observations	Key purpose: screening for psychopathy in youth 12-18 years No cut-off scores to make categorical classification of psychopathy or risk of violence. Total scores range from 0 to 40. Content: • narcissism • callous-unemotional • impulsivity	Predictive validity: • juvenile offenders in detention (n=117) retrospective prediction, APSD rating version AUC=.67; self-report version AUC=.68 (Douglas et al, 2008) Reliability: • adequate test re-test and internal consistency (see Murrie & Cornell, 2002). • correlations between parent and teacher ratings .26 and .40 (cited in Vincent, 2006)
PCS (Psychopathy Content Scale) - Murrie & Cornell (2000) State/country of origin: Canada Other versions: based on the 160-item Millon Adolescent Clinical Inventory MACI, Millon, 93) – a multiscale personality inventory Description: diagnostic screening Factors: personality Length: 25-items Completed by: young person self-report	Key purpose: Psychopathy – designed as a screening instrument to be followed by a more thorough assessment with PCL:YV Classification: Cut-off scores for high, medium, and low risk	Validity: • cut-off score of 11 - 85% of adolescent psychiatric inpatients correctly identified and 57% of low PCS scores. NB: 15% of highly psychopathic adolescents missed. (Murrie & Cornell, 2000) Reliability: • internal consistency =.87(Murrie & Cornell, 2000)

Table 5.2: More detailed assessment instruments

Details of instrument	Factors / content	Validity / reliability details
YLS/CMI (Youth Level of Service / Case Management Inventory) - Hoge & Andrews (2002) Country of origin: Canada Other versions: Adapted from LSI (Level of Supervision inventory) developed by Andrews (1982); Earlier youth versions YLSI (Andrews et al, 1984), LSI-R (Andrews & Bonta, 1995) Description: third generation, adjusted actuarial with professional override. Factors: static, dynamic Length: 42-items, 20-30 mins to complete Completed by: trained YJ practitioners or mental health professionals Sources of info: interview with youth, and other collateral info (e.g. file records)	Key purpose: case management, assessing risk, criminogenic need and responsivity Risk classification: 4 groups, low, moderate, high, and very high. Eight subscales: • prior and current offences • family circumstances/parenting • education/employment • peer assoc • substance abuse • leisure/recreation • personality/behaviour • attitudes/orientation Opportunity to indicate areas of strength is also provided	Predictive validity: • any recidivism across 22 different studies, mean weighted r=.32 (Olver et al, 2009) • across 11 studies, mean weighted AUC=.641 Schwalbe, 2007) Concurrent validity: • sig correlations with CBCL and YSR (Schmidt et al, 2005) Reliability: • inter-rater: YJ officers and MH professionals (n=29), ICCs r=.61 to .85 across sub-scales (Schmidt et al, 2005); ICC of 0.75 for total risk sore (Poluchowiz et al., 2000 cited in Vincent et al., 2009). • internal consistency: on 107 youth, Cronbach's alpha =.56 to .77 across sub-scales (Schmidt et al, 2005).
YLS/CMI-AA Australian Adaptation of YLS/CMI described above - Hoge & Andrews (1995) cited in Upperton & Thompson (2007) Work on this instrument began prior to the commercial availability of the parent instrument the YLS/CMI described above.	As above with the following adaptations: 1. 47-items – addition of 3 strength items (protective factors) 2. wording 3. tightening and printing of operational definition of items 4. reorganisation	Predictive validity: • any recidivism, 15m f-up, n=107: AUC=.75, r=.43 (Upperton & Thompson, 2007) • any recidivism, 6 to 32m f-up, n=174: AUC=.67, r=.28 (Thompson & Pope, 2005) Reliability: • test-retest stability (over 1-16m, mean=5m) r=.79 (Thompson & Pope, 2005)

Details of instrument	Factors / content	Validity / reliability details
ASSET – Youth Justice Board of England Wales (cited in Youth Justice Board, 2008) Country of origin: England and Wales Description: third generation. Factors: static and dynamic Length: 13 sections, 106-items, 4 hours to administer and score Completed by: trained YJ practitioners Sources of info: interview with youth, and other collateral info (e.g. file records)	Key purpose: case management, assessing risk, criminogenic need Risk classification: 5 groups, low, low-medium, medium, medium-high, high Domains: • individual • family • lifestyle • Community Sections: • offending career • living arrangements • family/personal relationships • education, training and employment • neighbourhood • lifestyle • substance abuse • physical health • emotional health • perception of self • thinking/behaviour • attitude to offending • motivation to change	Predictive validity: • any recidivism, 24m f-up, n=2,233): AUC=.731, (Baker et al, 2005) • any recidivism, 12m f-up, n=2,233): AUC=.719, (Baker et al, 2003) • predictive accuracy at 24m for sub groups (females offenders, ethnic minorities and younger offenders was maintained (Baker et al, 2005) Reliability: • inter-rater reliability: 60 different YJ practitioners ratings for 2 case studies, ICC =.23 to .52; % agreement 81-93%.

Details of instrument	Factors / content	Validity / reliability details
PCL-YV (Psychopathy Checklist - Youth Version) – Forth, Kosson, & Hare (2003), cited in Olver et al. (2009) Country of origin: Canada Other versions: Adapted from the adult PCL-R (Hare Psychopathy Checklist) first developed by Hare (1991) Description: diagnostic tool for psychopathy. Factors: static and dynamic Length: 20-items, 60-90 mins to complete Completed by: clinically trained expert raters Designed for: 12 to 17 years Sources of info: interview with youth and file analysis	Key purpose: a dimensional (level) assessment of the prototypical psychopath among adolescents. Defines psychopathy along interpersonal, affective and behavioural domains. No cut-off scores to make categorical classification of psychopathy or risk of violence. Total scores range from 0 to 40. Two factor scores for: • interpersonal affective dimensions; and • behavioural or lifestyle features	Predictive accuracy: • any recidivism across 20 different studies, mean weighted r=.28 (Olver et al, 2009) • across 3 studies, mean weighted AUC=.695 Schwalbe, 2007) Reliability: • inter-rater: excellent ICC .90-.92 across justice-involved institutions, probation and community samples (Forth et al., 2003 cited in Douglas et al., 2008) • internal consistency: Cronbach's alpha .85-.94 (Forth et al., 2003 cited in Douglas et al., 2008)
SAVRY (Structured Assessment of Violence Risk in Youth) - Borum, Bartel & Forth (2006) cited in Olver et al. (2009) Country of origin: United States Description: third generation / structured professional judgement Factors: static and dynamic Length: 30-items Completed by: qualified expert raters Designed for: 12 to 18 years Sources of info: interview with youth and file analysis	Key purpose: Assessment of risk of violence Risk classification: ratings of high, medium or low risk of violence, based on judgement (scores not added). Domains (24 risk factors, 6 protective factors): • historical including prior offending, poor school achievement (10-items), • individual (8-items) e.g. anger mgt, negative attitudes, low empathy • social/contextual (6-items) e.g. peer delinquency, lack of support • protective factors (6-items) e.g. prosocial involvement, strong social support	Predictive accuracy • any recidivism across 7 different studies, mean weighted r=.33 (Olver et al, 2009) • any recidivism, 12m f-up, ACU=.66 (Gretton & Abramowitz, cited in Vincent et al, 2009) Reliability: • inter-rater: acceptable ICC .77-.81 between students and experts (Catchpole and Gretton, 2003).

Details of instrument	Factors / content	Validity / reliability details
SECAPS (Secure Care Psychosocial Screening) – Putnins (1999) cited in Thompson & Putnins (2003) Country/state of origin: South Australia Description: third generation, risk, need, responsivity and risk of self-harming Factors: static and dynamic Length: 30-items, 30 mins to complete Completed by: trained YJ practitioners Sources of info: self-report via interview with youth	Key purpose: case management for youth placed in residential care. Content: • background (5-items) • sight and hearing (2-items) • natural family/residence (7-items) • literacy (2 tests) • numeracy (6-items) • nonverbal intelligence (1-item) • alcohol and drugs (13-items) • current mood (3-items) • self-destructiveness (5-items) • social relationships (3-items) • aggressiveness (2-items) • tattoos (2-items) • school/employment (6-items) • attention/restlessness (5-items)	Non located

Details of instrument	Factors / content	Validity / reliability details
VOINY (Victorian Offending Needs Indicatory for Youth), Department of Human Services, Youth Justice Development Unit, Victoria. (see Harris, 2008). Country/state of origin: South Australia Description: third generation Factors: static, dynamic (risk and protective factors) Length: 55-items (plus additional descriptive info) Completed by: trained YJ practitioners Sources of info: self-report via interview with youth	Key purpose: case management for youth given Youth Justice Orders. Classifies youth into four levels of need low, moderate, high, intensive. Provides information on level of intervention required to address identified criminogenic needs. Content: 1. Youth Offending • Offending profile (10 items) • Family circumstances (5 items) • Accommodation and finance (3 items) • Substance use (6 items) • Education, training and employment (5 items) • Peer relationships and community linkages (4 items) • Attitudes and behaviours (7 items) 2. Protective factors • Individual (7 items) • Family (5 items) • Community (3 items) 3. Special needs • Demographics • Health and development 4. Offence-related factors • Offence analysis • Attitudes	None located

5.3   Comparative performance of different instruments

Having presented details of the more widely used risk screening/assessment tools it is useful to consider what is known about their comparative effectiveness. A recent meta-analysis by Schwalbe (2007) identified 28 risk screening/assessment instruments that had been evaluated on 33 distinct samples.[18] Of these 16 were classified as second generation instruments based on actuarial modelling, with the remaining 12 being classified as third generation; including both risk and need factors.

Schwalbe carried out a meta-analysis of these 28 risk screening/assessment instruments to evaluate the ability of young offender risk instruments to predict recidivism. Overall, the 28 instruments were considered to have a medium performance level. The weighted average effect size across all instruments was AUC=.640 (equivalent to r=.24). This was lower, but not dissimilar, to that found with risk screening/assessment instruments for adult samples (r=.30).

Results of individual instruments revealed considerable variability with individual weighted effect sizes ranging from AUC .532 to .780. Those instruments achieving above AUC=.70 are considered acceptable (see Vincent et al., 2009), but those closer to .5 are achieving predictions little better than chance. Analysis by type of instruments suggested that briefer second generation risk screening/assessment instruments had smaller effect sizes (AUC=.635; r=.24) compared to the third generation instruments (AUC=.646; r=.26). No one instrument stood out as superior to others, the highest six effect sizes (all with AUCs above 0.7) are as follows:

- PCL-YV
- Child and Adolescent Functional Assessment Scale (CAFAS)
- YLS/CMI-AA, YLS/CMI
- Structured Assessment of Violence Risk in Youth (SARVY)
- ASSET.

Scwhalbe (2007) concluded that the findings provided support for the use of risk screening/assessment instruments to aid youth justice decision-making. However, because of the wide range of effect sizes they also suggested caution when selecting, implementing and interpreting individual risk screening/assessment instruments. If service planning is a priority then Schwalbe recommended instruments like YLS/CMI as being most useful as they measure criminogenic need that if targeted should reduce recidivism. However, if brevity and efficiency are a higher priority then locally derived actuarial risk screening/assessment instruments should be favoured.

Another recent and informative meta-analysis was carried out by Olver and colleagues (2009). They conducted a meta-analysis on what were considered to be the three most widely used tools (i.e. YLS/CMI; PCL-YV, and SAVRY). Each instrument predicted overall risk of recidivism but was designed with quite different objectives. The YLS/CMI assesses risk and criminogenic needs; the PCL-YV is a diagnostic tool for psychopathy and the SAVRY assesses risk of violence. Olver and colleagues identified 49 studies that used these tools. [19] They used meta-analysis techniques to evaluate the comparative performance of each instrument on a number of criteria, including their ability to predict general recidivism.

---

[18]   Follow-up times ranged from 6 to 60 months, although a 12m follow-up period was most common.

[19]   Twenty-two studies were identified that looked at the YLS/CMI or other youth versions of LSI; 28 that researched the PCL-YV; and 9 on the SAVRY. Mean follow-up time across studies was just over two years (28.4 months).

Similar to Schwalbe's (2007) findings, no single instrument was found to be superior, in regards to their ability to predict general recidivism. The SARVY and the YLS/CMI (or other youth versions of Level of Service Inventory [LSI])[20] both had mean weighted r=.32, with the PCL-YV achieving a slightly lower mean weighed r=.28. However, the authors noted there was large variability in magnitude of effect sizes for both YLS/CMI tools and the PCL-YV.

Olver et al. (2009) highlighted the importance for instruments to be able to generalise prediction across samples that are diverse with respect to gender, culture and ethnicity. Data to assess this were only available for YLS/CMI tools, but had positive results, continuing to significantly predict general recidivism among male, female, Aboriginal, and non-Aboriginal youth. Another meta-analysis by Schwalbe (2008) reviewed 19 studies that used a range of instruments to predict recidivism for both male and female young offenders and concluded that predictive validity estimates were equivalent for both.

Analysis of performance across different jurisdictions found that the YLS/CMI tools and the PCL:YV had significantly stronger predictive accuracy for Canadian studies compared to other jurisdictions. The authors suggested this could reflect that these instruments had been developed and normed using Canadian samples or could perhaps be explained by differences in recidivism as a result of differing approaches to youth justice of varying jurisdictions.

5.4   Summary

This chapter has provided details of youth risk screening/assessment tools that have been used in New Zealand and those from other jurisdictions which are accepted as having good validity and reliability. Information on the format and content of other risk screening/assessment instruments can be useful in compare that of the current YORST (e.g. predictive factors included, length, format, validity and reliability).

In New Zealand Department of Corrections has successfully integrated an actuarial model of risk prediction into their adult offender management system. It is a computerised generated score based on static factors. The youth justice system has lagged behind, there have been a couple of standardised risk screening/assessment tools developed (RSYO and YRS), but neither appear widely used and validation testing is incomplete. There appears to have been one attempt to develop an actuarial model for young offenders by researchers in Canterbury. Despite promising preliminary results this model appears to have received no further attention. Other screening of young offenders in New Zealand is that carried out by Child, Youth and Family social workers. Child, Youth and Family have put together a package of screening and assessment tools which has been successfully implemented nationally. This is currently under review with the aim of introducing tools with more established validity and reliability.

---

[20]   The YSL/CMI is a youth version of a risk assessment instrument designed for adults, the Level of Service Inventory (LSI).

Internationally in excess of 28 risk screening/assessment tools for young offenders have been developed. Just over half these were briefer actuarial tools, with the remaining being more detailed tools assessing both risk and needs. A meta-analysis has shown that overall these instruments are considered to have a medium performance in their ability to predict recidivism. Slightly lower, but not dissimilar, to that found with risk screening/assessment instruments for adults. No one instrument stood out as superior to others, but the highest effect sizes were found for the PCL:YV, CAFAS, YLS/CMI, SARVY and the ASSET. The YLS/CMI has the strongest evidence of being valid across gender and Aboriginal and non-Aboriginal youth and across different jurisdictions.

There are a range of valid and reliable instruments developed to predict risk of recidivism. Reviews of these instruments suggest if service planning is a priority than instruments like the YLS/CMI that measure criminogenic need are likely to be most useful. However, if brevity and efficiency are a higher priority then locally developed actuarial models should be considered.

6   Implications for YORST

The key purpose of this literature review was to inform a larger programme of research which aims to assess the validity and reliability of the New Zealand Police YORST. The discussion below is not meant to form a conclusive appraisal of the YORST but rather to place the findings if this review into context and highlight areas requiring closer examination in the research to follow.

6.1   Purpose and type of instrument

Development: The YORST has been adapted from an earlier risk screening tool known as the ARNI. Both tools have been developed based largely on practitioner input (experienced Police Youth Aid Officers and Police Youth Development staff). The YORST has been further developed based on a review and consideration of empirically-based risk factors. Individual items of the YORST have not been subjected to statistical analysis of their association with recidivism.

Type: According to Schwalbe's (2007) classification, the YORST is best described as a third generation instrument as it includes both static and dynamic factors and aims to inform case management in addition to estimating risk of recidivism. Of note, some items on the YORST that are described as dynamic appear to differ from the usual definition of dynamic (e.g. behaviour or attitudes amenable to change). For example, question 4 focuses on the seriousness of the current offence and this question is marked as a dynamic factor. This area of questioning is more commonly reported as a static factor in comparable instruments and it would appear that the YORST uses dynamic to relate to factors that can change over time rather than if the variable is amenable to intervention.

There is some rater or examiner discretion in the YORST, which the research suggests is a positive characteristic as this provides the rater with the ability to account for idiosyncratic risk factors (e.g. individual characteristics / circumstances).

Length: The YORST is longer than many of the brief second generation instruments that primarily aim to predict risk. Several of these instruments accurately predict recidivism with five or fewer items (e.g. SECAPS Recidivism Risk Index). However, the YORST is considerably shorter than other third generation tools which, in addition to predicting risk, aim to assess criminogenic needs in order to inform case management. The brevity of the YORST means it can be applied quickly to a large number of young offenders, and therefore meets its objective of being a screening tool for Police Youth Aid Officers. However, careful consideration will need to be given to how this screening information is used to influence decision-making if further more detailed assessment is not carried out.

Decision point and information requirements: The YORST has been designed to be used at the point when children (10-13 years) and youth (14-16 years) come to police notice due to alleged offending. This represents a large group of young offenders the majority of whom will be low risk offenders who are typically dealt with through diversionary measures such as a warning or alternative action. A smaller proportion of this group would be considered high-risk. The minimum criteria for the completion of YORST prioritises its use with those likely to be higher risk young offenders:

- all child offenders (10-13 years) at the second offence

- all child offenders (10-13 years) referred to Child Youth and Family for care and protection due to serious offending

- all youth offenders (14-15 years) being referred for Youth Justice Family Group Conferences (including arrest).

The YORST must therefore be capable of accurately differentiating the level of risk among this particular group of young offenders, and in particular to identify those who are at highest risk of being serious and chronic offenders.

The intended purpose of the YORST is to identify a young persons risk of re-offending and to provide the foundation for a targeted and appropriate response. It is also hoped the YORST will be able to measure change in level of risk, thus providing a measure of the impact of police interventions. Despite being applied at one decision point, a large range of information is required from the YORST, but must be completed in a time efficient manner. These competing requirement of the YORST may in part reflect the nature of police work in New Zealand, where compared to other jurisdictions officers have a high degree of autonomy in deciding what level of action will be taken for any young offender they come into contact with.

The inclusion of dynamic factors in the YORST therefore fits with its stated objectives, and those of the Children, Young Persons and Their Families Act 1989, however, careful attention needs to be given to whether the number of dynamic items on the YORST are sufficient to inform case management and detect change in risk over time.

An important purpose of standardised assessment is allowing for consistency in youth justice decision-making. The YORST clearly meets this objective and, providing there is a sufficiently good uptake of the tool, will also allow regional comparisons to be made and ensure the availability of data for further empirical research on New Zealand young offenders. Another positive attribute of the YORST is the existence of a policy document, the presence of which is identified in the literature as a key element for a standardised assessment process.

6.2   Scoring

The scoring system for the YORST appears to be based on the cumulative risk model or Burgess method, where item responses are summed to produce a risk score. Despite this simple approach to scoring, research has shown this method to be equivalent and sometimes superior to other more sophisticated statistical methods. Later analysis of YORST data would be able to assess the comparative value of different scoring methods.

6.3   Inclusion of predictive factors

Chapter two reviewed factors identified in the literature that are associated with recidivism in young offenders. Perhaps not surprisingly, all items on the YORST, appear in the comprehensive list of factors listed, or represent a proximal estimate of factors in Table 1. For example, question 1 on the YORST 'time since last came to police notice' could be considered a proximal estimate for number of prior offences (i.e. a higher frequency of offending would likely be associated with a short time period between previous contacts).

_____

All items on the YORST also appear in Table 2, however, some such as those related to parent pathology (Question 13 & Question 14) and school attendance (Question 8) were found by Cottle et al. (2001) to be non-significant predictors. However, other researchers have found these to be significant (e.g. Fergusson & Horward, 2002; Maxwell et al., 2004; Moffitt & Caspi, 2001). Some of the factors that were significant predictors in Table 2 appear not be included in the YORST, these include:

• age at first commitment (placement in a correctional facility)

• non-severe pathology

• conduct problems

• effective use of leisure time

• length of first incarceration

• number of prior commitments.

It should be noted that factors listed above are based on international research and factors related to youth incarceration (or commitment) are less relevant in New Zealand with a very small number of young offenders being exposed to this type of intervention. It must also be remembered that it must be feasible to complete an instrument with available resources and types of skills. For instance assessment of psychopathology may not be considered appropriate for frontline professionals such as police officers.

Research carried out in New Zealand found the frequency of previous contact with police and/or CYF has been found to be a significant predictor of recidivism in New Zealand young offenders. This information is captured by a number of YORST items. Question 9 of the YORST which refers to prior contacts with CYF. This captures relevant factors such previous exposure to foster/state care, abuse, poor family functioning. It is also taps into early offending, as in New Zealand child offenders are dealt with differently to many other jurisdictions, in that in the main, any offending by individuals under the age of 14 years is dealt with as a care and protection issue rather than a youth justice matter. Hence, referral to CYF is also a measure of early offending. Other questions in the YORST are proximal factors for previous contact with police (e.g. Questions 1 and 2), but other measures specifically relating to frequency of contact with police may also be worth considering.

A finding of this review was the importance for a screening/assessment tool to be valid across different sub-groups of offenders including gender, ethnicity and age. There were no clear guidelines on how instruments should be adapted for different groups, however, it will be important with later testing of the YORST to assess its validity across sub-groups of young offenders. Its applicability to the child offenders (those aged 10-13years) will require particular attention.

6.4   Future evaluation activities

Research to evaluate the YORST is in its early stages. This review has presented details of other established risk screening/assessment instruments which provide a useful basis to compare the format and content of the YORST (e.g. factors included, length, format, validity and reliability). Subsequent evaluation activities will include an analysis of the accuracy and completeness of YORST data being collected, followed by testing of its validity and reliability. This review has highlighted the need for careful consideration in assessing the validity of the YORST across gender, ethnicity and age bands. It may also be useful to monitor the uptake of the tool and evaluate the policies, protocols and training that impact on how the tool is implemented.

References

Andrews, D. A., Bonta, J. and Hoge, R. D. (2000). 'Classification for effective rehabilitation: rediscovering psychology'. Criminal Justice and Behavior, 17: 19-52.

Andrews, D. A., Bonta, J. and Wormith, J. S. (2006). 'The Recent Past and Near Future of Risk and/or Need Assessment'. Crime Delinquency, 52 (1): 7-27.

Arnull, E., Eagle, S., Gammampila, A., Archer, D., Johnston. V., Miller, K and Pitcher, J. (2005). Persistent Youth Offenders: A retrospective study. Youth Justice Board for England and Wales 2005.

Bailey, S., Doreleijers, T. and Tarbuck, P. (2006). 'Recent developments in mental health screening and assessment in juvenile justice systems'. Child and Adolescent Psychiatric Clinics of North America, 15 (2): 391-+.

Baker, K., Jones, S., Merrington, S. and Roberts, C. (2005). Further Development of Asset. London: Youth Justice Board for England and Wales.

Baker, K., Jones, S., Roberts, C. and Merrington, S. (2003). ASSET: The Evaluation of the Validity and Reliability of Youth Justice Board's Assessment for Young Offenders. London: Youth Justice Board for England and Wales.

Bakker, L., O'Malley, J. and Riley, D. (1998). Storm warning: Statistical models for predicting violence. Christchurch: Department of Corrections.

Bakker, L., O'Malley, J. and Riley, D. (1999). Risk of reconviction: Statistical models predicting four types of re-offending. Christchurch: Department of Corrections.

Becroft, A. (2004). Youth Justice in New Zealand: Future Challenges. New Zealand Youth Justice Conference: "Never Too Early, Never Too Late." Wellington, 17-19 May 2004.

Cain, M. (1996). Recidivism in Juvenile Offenders in New South Wales. Haymarket, New South Wales: Department of Juvenile Justice.

Catchpole, R. E. H. and Gretton, H. M. (2003). 'The predictive validity of risk assessment with violent young offenders - A 1-year examination of criminal outcome'. Criminal Justice and Behavior, 30 (6): 688-708.

Child Youth and Family. (2006). Risk and Needs Assessment. Presentation, Supported Bail National Hui, April 2006, Wellington.

Cottle, C. C., Lee, R. J. and Heilbrun, K. (2001). 'The prediction of criminal recidivism in juveniles - A meta-analysis'. Criminal Justice and Behavior, 28 (3): 367-394.

Crawford, R. and Kennedy, P. (2008). Improving Interventions to Reduce Violent Offending by Young People in New Zealand. August 2008. Wellington: Ministry of Justice.

Cunneen, C. (2005). Racism, discrimination and the over-representation of indigenous people in the criminal justice system: Some conceptual and explanatory Issues. Institute of Criminology, University of Sydney Law Faculty, Paper presented to the ANZSOC Conference, Wellington, New Zealand.

Domburgh, L. V., Loeber, R., Bezemer, D., Stallings, R. and Stouthamer-Loeber, M. (2009). 'Childhood predictors of desistance and level of persistence in offending in early onset offenders'. Journal of Abnormal Child Psychology, 37 (7): 967-980.

Douglas, K. S., Epstein, M. E. and Poythress, N. G. (2008). 'Criminal recidivism among juvenile offenders: Testing the incremental and predictive validity of three measures of psychopathic features'. Law and Human Behavior, 32 (5): 423-438.

Duncan, A. (2009). Child and Youth Offending Statistics in New Zealand: 1992 to 2007. Wellington: Ministry of Justice.

Fergusson DM, Horwood LJ, Lynskey MT. Ethnicity and bias in police contact statistics. Australian and New Zealand Journal of Criminology, 1993; 26: 193-206.

Fergusson, D. M. and Horwood, J. L. (2002). 'Male and female offending trajectories'. Development and Psychopathology, (14): 159-177.

Fergusson, D. M., Horwood, J. L. and Nagin, D. S. (2000). 'Offending trajectories in a New Zealand birth cohort'. Criminology, 38: 401-427.

Fergusson DM. Ethnicity and interpersonal violence in a New Zealand birth cohort. In

Hawkins, Darnell F. (Ed). Violent Crimes: Assessing Race and Ethnic Differences. Cambridge: Cambridge University Press, 2003, pp 138-153.

Galletly, A. (2006). Frequency and severity of offending young people in New Zealand: Descriptive analysis and development. Unpublished Masters, Canterbury University, Christchurch.

Gavazzi, S. M., Yarcheck, C. M., Sullivan, J. M., Jones, S. C. and Khurana, A. (2008). 'Global risk factors and the prediction of recidivism rates in a sample of first-time misdemeanant offenders'. International Journal of Offender Therapy and Comparative Criminology, 52 (3): 330-345.

Gottredson, D. M. and Snyder, H. N. (2005). The Mathematics of Risk Classification: Changing data into valid instruments of juvenile courts. Washington, D.C.: National Centre of Juvenile Justice. Department of Justice. Office of Juvenile Justice and Delinquency Prevention.

Grace, R., McLean, A. and Warren, J. (2006). An Evaluation of the Reducing Youth Offending Programme. (Unpublished report for Child Youth and Family Services, Wellington. Prepared by University of Canterbury (Canterprise) and Centre for Research, Evaluation and Social Assessment (CRESA)). April 2006.

Hanson, R. K. and Morton-Bourgon, K. E. (2009). 'The Accuracy of Recidivism Risk Assessments for Sexual Offenders: A Meta-Analysis of 118 Prediction Studies'. Psychological Assessment, 21 (1): 1-21.

Harris, R. (2008). Risk/Needs assessment and response to criminogenic factors with young people. Young People, Crime and Community Safety: engagement and early intervention. 25-26 February, 2008, Melbourne, Australia

Hawkins, J. D., Herrenkohl, T. I., Farrington, D. P., Brewer, D., Catalano, R. F., Harachi, T. W. and Cothern, L. (2000). Predictors of youth violence.

Hoge, R. D. (2002). 'Standardized instruments for assessing risk and need in youthful offenders'. Criminal Justice and Behavior, 29 (4): 380-396.

Hoge, R., and Andrews, D.A. (2002), YLS/CMI Youth Level of Service/Case Management Inventory Users Manual. Toronto: Mental Health Systems.

Holtfreter, K. and Cupp, R. (2007). 'Gender and risk assessment: The empirical status of the LSI-R for women.' Journal of Contemporary Criminal Justice 23 (4): 363-382.

Howell, J. (1995). Guide for Implementing the Comprehensive Strategy for Serious, Violent, and Chronic Juvenile Offenders. Washington, DC: Office of Juvenile Justice and Delinquency Prevention, U.S. Department of Justice.

Johansson, P. and Kempf-Leonard, K. (2009). 'A gender-specific pathway to serious, violent and chronic offending?: Exploring Howell's risk factors for serious delinquency'. Crime and Delinquency, 55: 216-239.

Kalb, L. M. (2006). Screening for Violent Recidivism in the Juvenile Justice System: Establishing accuracy. Unpublished Doctoral Thesis, University of Pittsburgh, Pittsburgh.

Krysik, J. and LeCroy, C. W. (2002). 'The empirical validation of an instrument to predict risk of recidivism among juvenile offenders'. Research on Social Work Practice, 12 (1): 71-81.

Maxwell, G., Kingi, V., Robertson, J., Morris, A., Cunningham, c. (2004). Achieving Effective Outcomes in Youth Justice Final Report. Ministry of Social Development.

Maxwell, G. and Morris, A. (1999). Understanding reoffending. Wellington: Institute of Criminology, Victoria University of Wellington.

Maxwell, G. and Robertson, J. (1995). Child Offenders. Wellington: Office of the Commissioner for Children.

McLaren, K. (2000). Tough is Not Enough - Getting Smart About Youth Crime. Wellington: Ministry of Youth Affairs.

McLaren, K. (2009). Alternative Actions that Work: a review of the research on police warnings and alternative action. Unpublished draft report, Youth Services Group, New Zealand Police, Wellington.

Miller, J. and Lin, J. (2007). 'Applying a generic juvenile risk assessment instrument to a local context - Some practical and theoretical lessons'. Crime & Delinquency, 53 (4): 552-580.

Ministry of Justice. (2002). Youth Justice and Social Sector Service Delivery to the Children and Young People Convicted in relation to the death of Michael Choy. November 2002. Wellington: Ministry of Justice.

Ministry of Justice and Development, M. o. S. (2002a). Report of the Ministerial Taskforce on Youth Offending. April 2002. Wellington: Ministry of Justice and Ministry of Social Development.

Ministry of Justice and Development, M. o. S. (2002b). Youth Offending Strategy: Preventing and reducing offending and re-offending by children and young people. April 2002. Wellington: Ministry of Justice and Ministry of Social Development.

Moffitt, T. E. (1993a). 'Adolescence-limited and life-course-persistent behaviour - A developmental taxonomy". Psychological Review, 100 (4): 674-701.

Moffitt, T. E. (1993b). 'Adolescence-limited and life-course-persistent antisocial behaviour - A developmental taxonomy'. Psychological Review, 100 (4): 674-701.

Moffitt, T. E. and Caspi, A. (2001). 'Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females'. Development and Psychopathology, 13 (2): 355-375.

Murrie, D. C. and Cornell, D. G. (2002). 'Psychopathy screening of incarcerated juveniles: A comparison of measures'. Psychological Assessment, 14 (4): 390-396.

New Zealand Police. (2009). Request for Proposal: Research to Validate the New Zealand Police Youth Offending Risk Screening Tool (YORST) - First year Wellington: New Zealand Police.

O'Mahoney, P. (2009). 'The risk factors prevention paradigm and the causes of youth crime: A deceptively useful analysis?' Youth Justice, 9 (2): 99-114.

Odgers, C. L., Moffitt, T. E., Broadbent, J. M., Dickson, N., Hancox, R. J., Harrington, H., Poulton, R., Sears, M. R., Thomson, W. M. and Caspi, A. (2008). 'Female and male antisocial trajectories: From childhood origins to adult outcomes'. Development and Psychopathology, 20 (2): 673-716.

Olver, M. E., Stockdale, K. C. and Wormith, S. (2009). 'Risk assessment with young offenders: A meta-analysis of three assessment measures'. Criminal Justice and Behaviour, 36: 329-353.

Putnins, A. L. (2005). 'Assessing recidivism risk among young offenders'. Australian and New Zealand Journal of Criminology, 38 (3): 324-339.

Roguski, M. (2009). Te Whare Ruruhau o Meri. Wellington: Te Puni Kokiri.

Rutter, M., Giller, H., & Hagell, A. (1998). Antisocial behaviour by young people. Cambridge, United Kingdom: Cambridge University Press

Schmidt, F., Hoge, R. D. and Gomes, L. (2005). 'Reliability and Validity Analyses of the Youth Level of Service/Case Management Inventory'. Criminal Justice and Behavior, 32 (3): 329-344.

Schwalbe, C., S. (2009). 'Reducing risk for violence and aggression in youth'. In, J. Andrade (Ed.), Handbook of violence risk assessment and treatment: new approaches for mental health professionals.

Schwalbe, C. S. (2007). 'Risk assessment for juvenile justice: A meta-analysis'. Law and Human Behavior, 31 (5): 449-462.

Schwalbe, C. S. (2008a). A meta-analysis of juvenile justice risk assessment instruments: Predictive Validity by Gender'. Criminal Justice and Behavior, 35 (11): 1367-1381.

Schwalbe, C. S. (2008b). 'A meta-analysis of juvenile justice risk assessment instruments: Predictive validity by gender'. Criminal Justice and Behavior, 35 (11): 1367-1381.

Schwalbe, C. S., Fraser, M. W., Day, S. H. and Arnold, E. M. (2004). 'North Carolina Assessment of Risk (NCAR): Reliability and Predictive Validity with Juvenile Offenders'. Journal of Offender Rehabilitation, 40: 1-22.

Schwalbe, C. S., Fraser, M. W., Day, S. H. and Cooley, V. (2006). 'Classifying juvenile offenders according to risk of recidivism - Predictive validity, race/ethnicity, and gender'. Criminal Justice and Behavior, 33 (3): 305-324.

Shader, M. (2003). Risk factors for delinquency: An overview. Office of Juvenile Justice and Delinquincy Prevention: US Department of Justice

Smith, W. and Aloisi, M. (1999). 'Prediction of recidivism among "second timers" in the juvenile justice system: Efficiency of screening chronic offenders'. American Journal of Criminal Justice, 23 (2): 201-222.

Stoolmiller, M. and Blechman, E. A. (2005). 'Substance use is a robust predictor of adolescent recidivism'. Criminal Justice and Behavior, 32 (3): 302-328.

Te Puni Kokiri (2000). Whanake Rangatahi: Programmes and services to address Māori youth offending. Wellington: Te Puni Kokiri.

The Werry Centre. (2009). A Literature Review: Mental health and alcohol and other drug screening, assessment and treatment for youth justice populations. Auckland: The Werry Centre for Child and Adolescent Mental Health Workforce Development.

Thompson, A. P. (2003). Current research and practice with risk-need assessments in NSW juvenile justice. Juvenile Justice: From Lessons of the Past to a Road Map for the Future, Sydney, 1-2 December 2003.

Thompson, A. P. and Pope, Z. (2005). 'Assessing juvenile offenders: Preliminary data for the Australian Adaptation of the Youth Level of Service/Case Management Inventory (Hoge & Andrews, 1995)'. Australian Psychologist, 40 (3): 207-214.

Thompson, A. P. and Putnins, A. L. (2003). 'Risk-need assessment inventories for juvenile offenders in Australia'. Psychiatry, Psychology and Law: 324-333.

Trochim, W.M. (2006). Types of reliability, Research Methods Knowledge Base, accessed 4 June 2010 http://www.socialresearchmethods.net/kb/reltypes.php

Upperton, R. A. and Thompson, A. P. (2007). 'Predicting juvenile offender recidivism: Risk-need assessment and juvenile justice officers'. Psychiatry Psychology and Law, 14 (1): 138-146.

van der Laan, A., Blom, M. and Kleemans, E. (2009). 'Exploring long-term and short-term risk factors for serious delinquency'. European Journal of Criminology, 6 (5): 419-438.

Vincent, G., Terry, A. and Many, S. (2009). 'Chapter 12: Risk/Needs Tools of Antisocial Behavior and Violence Among Youthful Populations'. In, J. T. Andrade (Ed.), Handbook of Violence Risk Assessment and Treatment: New Approaches for Mental Health Professionals. New York: Springer Publishing Company, LLC.

Vincent, G. M. (2006). 'Psychopathy and violence risk assessment in youth'. Child and Adolescent Psychiatric Clinics of North America, 15 (2): 407-+.

Watt, B., Howells, K. and Delfabbro, P. (2004). 'Juvenile recidivism: Criminal propensity, social control and social learning theories'. Psychiatry, Psychology and Law, 11 (1): 141:153.

Weatherburn, D., Cush, R. and Saunders, P. (2007). 'Screening juvenile offenders for further assessment and intervention'. Crime and Justice Bulletin, (109): 1:11.

Webster, C., MacDonald, R. and Simpson, M. (2006). 'Predicting Criminality? Risk Factors, Neighbourhood Influence and Desistance'. Youth Justice, 6 (1): 7-22.

Whiting, E. and Cuppleditch, L. (2006). Re-offending of Juveniles: results from the 2004 cohort. Home Office Statistical Bulleting. London: Home Office.

_____

Wilson, N. and Rolleston, R. (2004). A Risk-Need Profile Using Four Measures for Youth Offenders Incarcerated in Young Offender Units. Retrieved 7 December, 2009, from http://www.corrections.govt.nz/research/a-risk-need-profile-youth-offenders.html

Youth Justice Board. (2005). Risk and Protective factors. London: Youth Justice Board for England and Wales.

Youth Justice Board. (2008). Assessment, planning interventions and supervision: Source document. London: Youth Justice Board for England and Wales.

Appendix 1: The YORST

YOUTH OFFENDING RISK SCREENING TOOL
New Zealand POLICE — Nga Pirihimana O Aotearoa

NAME (Child/YP):	Surname	First name(s)	NIA Person ID No:	
			File no:	
DOB		Age	Gender	Male / Female	Date RST Completed	by (QID)

ETHNICITY	European	Pacific	Asian	Other
	Maori	Iwi		Hapu

| Incident / Offence Code | | Incident / Offence Description | |

Part (A) Offending Factors

1. Time since last came to Police notice for their offending?

No previous	Over 2 yrs	1 to 2 yrs	Less than 1 yr	1 to 6 mths	Under 1 mth	D
0	1	2	3	4	5	

2. Time since last came to Police notice for incidents (e.g. 1J, 2M, 1T) relating to them and/or serious behaviour incident at school?

No previous	Over 2 yrs	1 to 2 yrs	Less than 1 yr	1 to 6 mths	Under 1 mth	D
0	1	2	3	4	5	

3. Highest level of previous intervention? (final outcome)

No previous	Noting	Warning	Alt. Action	FGC	Youth Court
0	1	2	3	4	5

4. At what age was offending first reported to Police (if first offence use current age)?

No offences	15+	14	13	10 to 12	Under 10
0	1	2	3	4	5

5. Rate the seriousness of the current primary offence using the youth offence rating tool (see A4 list).

Minimum	Minimum / Medium	Medium	Medium / Maximum	Maximum	D
1	2	3	4	5	

6. Is the nature (MO) of current or previous offending of a concerning nature?

Very Low	Low	Medium	High	Extreme	D
1	2	3	4	5	

| Comments re Question 6: | |

Part (B) Peer Group Factors

7. Influential peers known to Police?

None	Very few known	Some known	Many known	All known repeat offenders	Unknown	D
0	1	3	4	5	0	

Part (C) Education / Employment Factors (contact the school, but not the employer)

8. Current school / education / course or employment status

Full time well engaged	Full time some issues	Mostly attends	Irregular attendance	Stood down / suspended	Not attending (school / job)	Unknown	D
0	1	2	3	4	5	0	

Part (D) Care & Protection History

9. Has a notification been made to CYF for this family or child / young person?

No	Notification concerning another sibling	Notification concerning this child / young person	Some form of intervention provided by Child, Youth & Family	Currently / previously in the custody of CYF (101 status)
0	2	3	4	5

Part (E) Alcohol and/or Drug Use

10. Is their use of alcohol or drugs causing concern? (consider the long term effects of the type of drugs used).

No concern	Slight	Moderate	Serious	Very Serious	Unknown	D
0	1	2	4	5	0	

Part (F) Family Factors

If there are FAMILY VIOLENCE records in NIA for this family / address, what is the highest FV score?

11	Zero Records	Records, but no score	Score from 1 - 8	Score from 9 - 16	Score 17 or over
	0	2	3	4	5

Where do they live? (socio economic area decile rating of local state primary school)

12	8 - 10	4 - 7	2 - 3	1	Transient / Motor Camp	D
	0	2	3	4	5	

Are there concerns in the living situation? e.g. parent / caregiver support and supervision of child / young person, parental mental health problems, drug and alcohol use, suspected child abuse and / or unrecorded family violence

13	None	Very minor concerns	Some concerns	Major concerns	Some major concerns	Young Person Transient	Unknown	D
	0	1	2	3	4	5	0	

Detail Concerns:	

Family members have offending history?

14	None	Parent(s) with minor history	Parent/s with major history (imprisonment)	Parent(s) have offended within past 12 months	Sibling(s) have offended within last 12 months	Unknown
	0	2	3	4	5	0

Any General Comments:	

Information Sources

Spoken To		Child / young person	Parent / caregiver	School / course provider / MOE	Child Youth & Family	Other agency
	This time					
	Previously					
	Not At All					

Scoring Instructions / Risk Screening

Questions			Answers		YORST Score	
No. of Questions		Max	Sum of the Scores (Above)	=	x 100 =	%
Not Answered:	Answered:	x 5	Max. Total for Answered Questions	=		

Dynamic Risk Factors

Static Factor Results		Dynamic YORST Score	
Sum of Dynamic Factors		x 100 =	%
Maximum Possible Total for Dynamic Factors	45		

Youth Aid Response / Your Station

Warning	AA	FGC	Youth Court	Police Youth Development	Other	

Appendix 2: Methodological approach of the literature review

The first stage to the literature review was a systematic search of existing information sources. This included relevant information already held by NZ Police and also that held by CJRC and the Institute of Criminology.

Our search then continued to cover academic databases of multidisciplinary journals (e.g. criminology, psychology, sociology and gender studies). Databases searched included - Web of Science, PSYCH Info, ProQuest, Sociological Abstracts and Te Puna (Index of New Zealand).

Each database has Key descriptive terms to cover areas of literature, those relevant to the current review were identified and combined appropriately (e.g. Web of Knowledge –'Juvenile-offenders'; 'Predictive-validity'; 'recidivism'; 'Assessment').

We also conducted extensive internet searches using google and google scholar, and a search of specific websites of government, professional and other organisations who produce criminological research, including:

- Australian Institute of Criminology

- British Home Office

- Youth Justice Board

- National Institute of Justice

- Campbell Collaboration Crime and Justice Group

- National Criminal Justice Reference Service

- Office of Juvenile Delinquency Prevention Programme.

All located literature was entered into *Endnote* (bibliographic software) to assist with the management, organisation and later referencing of research articles.

Appendix 3: Possible predictor variables

Table A1:   Predictor varibles considered for inclusion in RYOP model

CYRAS predictor variables	NIA Predictor variables
• **MaleFemale** • **MãoriPacific** • **AgeAtFirstIntake** • AgeAtCriterion • **AgeAtFirstYJ** • **NumPriorCP** • **NumPriorYJ** • NumPriorIntakes • NumPriorIntakesSec15 • NumPriorIntakesUrgent • NumPriorIntakes>Age1 • NumPriorPlacements • NumFindings • NumFindingsBRDifficulties • NumFindingsEmotAbuse • NumFindingsNeglected • NumFindingsNotFound • NumFindingsPhysicalAbuse • NumFindingsSelfHarm • **NumFindingsSexualAbuse** • NumPriorYJ-FGC • NumPriorCourtOutcomes • NumPSupervisionOutcomes • **NumPCourtDates** • NumPCustodyOutcomes • NumPOtherYJOutcomes • NumPriorYJ-FGCNoAgree • YJPost	• NumPriorProsecutions • NumPriorRecords • NumPriorChargeSheets • **NumPriorIntelligence** • **NumPriorOccurrence** • NumPriorYouthAid • NumProsecFollowUp • MaxSeverity • TotalSeverity

Those included in predictive models are bolded (see Galletly, 2006).