

Commissioned by  
New Zealand Police

Research to validate the New Zealand  
Police Youth Offending Risk Screening Tool  
(YORST) – Phase III

FINAL REPORT

by  
Dr Elaine Mossman

Final Report  
31 May 2016

# Contents

<b>Acknowledgements.....</b>	<b>i</b>
<b>Executive Summary .....</b>	<b>ii</b>
<b>1 Introduction .....</b>	<b>1</b>
1.1 Background .....	1
1.2 Key findings from Phases I and II .....	2
1.2.1 Phase I (a) - Literature review .....	2
1.2.2 Phase I (b) - Analysis of the quality of initial YORST data .....	2
1.2.3 Phase II – Predictive validity of the initial YORST (version 1) .....	3
1.3 Objectives of Phase III .....	4
<b>2 Methodology.....</b>	<b>5</b>
2.1 Evaluation objectives .....	5
2.2 Details of evaluation methodology .....	5
2.2.1 Concurrent validity and inter-rater reliability of finalized YORST(v2) .....	5
2.2.2 Predictive ability of finalised YORST(v2) and the Mini YORST .....	6
<b>3 Inter-rater reliability and concurrent validity.....</b>	<b>10</b>
3.1 Inter-rater reliability .....	10
3.1.1 Inter-rater reliability of YORST(v2) Total risk score.....	10
3.1.2 Inter-rater reliability of YORST(v2) individual items.....	11
3.2 Concurrent validity .....	13
3.2.1 Procedure .....	14
3.2.2 Total risk ratings.....	15
3.3 Summary.....	16
<b>4 Predictive ability.....</b>	<b>18</b>
4.1 Sample characteristics .....	18
4.2 Percentage of offenders correctly predicted .....	19
4.3 Accuracy of the prediction across the range of total risk scores.....	20
4.4 Survival analysis .....	22
4.5 Mean YORST(v2) scores of those re-apprehended compared to those not .....	23
4.6 Point bi-serial correlation.....	23
4.7 Receiver Operating Characteristics (ROC) analysis.....	24
4.8 Relationship between YORST and other measures of recidivism .....	26
4.9 Accuracy of the Mini YORST .....	27
4.10 Summary.....	28
<b>5 Item response analysis.....</b>	<b>30</b>
5.1 Item analysis – frequency of responses.....	30
5.2 Inter-correlation between items.....	30

5.3	Logistical regression .....	32
5.4	Mini YORST weightings .....	33
5.5	Alternative predictor items.....	33
5.6	Summary.....	35
<b>6</b>	<b>Conclusion .....</b>	<b>37</b>
<b>7</b>	<b>References.....</b>	<b>38</b>
	<b>Appendix 1: YORST (v2) .....</b>	<b>39</b>
	<b>Appendix 2: Case study vignette .....</b>	<b>43</b>

## Tables

Table 3.1	Inter-rater reliability on individual YORST items .....	12
Table 3.2	Comparison of content of YORST and YLS/CMI 2.0 .....	14
Table 3.3	Correlation of YORST and YLS/CMI 2.0 (n=28) .....	16
Table 4.1	Demographics of research sample .....	18
Table 4.2	Percentage of cases correctly predicted at 12 months .....	20
Table 4.3	Percentage of cases correctly predicted at 24 months .....	20
Table 4.4	Percentage of cases re-apprehended across low, medium and high YORST(v2) total risk scores .....	21
Table 4.5	Difference in mean total YORST risk scores of those re-apprehended and those not re-apprehended .....	23
Table 4.6	ROC analyses for YORST (v2) total risk scores .....	26
Table 4.7	ROC analyses for Mini YORST .....	28
Table 5.1	Correlations with YORST(v2) items 1 to 14 with each other and recidivism (12m and 24m).....	31
Table 5.2	Logistic Regression analysis of YORST(v2) items .....	32
Table 5.3	Correlations of incident variables (total number) with recidivism (12m and 24m).....	34
Table 5.4	Logistic Regression analysis of YORST(v2) items and possible alternative predictors .....	35

## Figures

Figure 3.1	Distribution of YORST total risk scores among 72 raters .....	10
Figure 3.2	Distribution of total risk scores .....	15
Figure 4.1	YORST(v2) total risk scores of sample (n=1212) .....	19
Figure 4.2	YORST(v2) scores against percentage re-apprehended .....	21
Figure 4.3	Kaplan-Meier Survival Function for 24 month recidivism for young offenders classified as low, medium or high risk according to YORST(v2).22	
Figure 4.4	ROC curve for the YORST(v2) total risk score over 12 months (n=1212); AUC=.759 .....	24
Figure 4.5	ROC curve for the YORST(v2) total risk score over 24 months (n=1212); AUC=.774 .....	25



# Acknowledgements

Anne Harland of New Zealand Police project managed this final phase of research and provided valuable support and feedback, whilst Glen Morrison had the vital role of extracting and sourcing the data required for this research, with assistance from Senior Sergeant Greg Clark.

Independent reviews and valuable feedback in relation to this report were provided by Su-Wuen Ong from the Department of Corrections and Associate Professor Anthony Thompson and Dr Andrew McGrath of Charles Sturt University in New South Wales.

A final acknowledgement needs to be given to the members of the New Zealand Police Youth Services team for their commitment over the years to the development of an evidence-based risk assessment tool. This work began in 2003 with the development of Adolescent Risk Needs Inventory (ARNI). Melanie Atkinson was set the task of reviewing and further developing this tool and, with the assistance of Christine Jamieson, this resulted in the introduction of the first version of the Youth Offending Risk Screening Tool (YORST) in 2007.

# Executive Summary

This report presents the findings from the final phase of research to assess the validity, reliability and predictive capability of New Zealand Police's Youth Offending Risk Screening Tool (YORST).

Standardised risk and needs assessment and/or screening of young offenders are recognised internationally as best practice. New Zealand Police began work to develop a risk screening tool in 2003, resulting in the development and piloting of the YORST in 2007. The YORST was rolled out nationally in July 2009 following a favourable evaluation of the pilot.

A risk screening/assessment instrument that improves the quality of youth justice decision-making must be both valid and reliable as well as user friendly. Hence, a programme of research was initiated in 2009 to evaluate the performance of YORST against these psychometric properties.. Following promising results from phase I and II, the YORST was revised to maximise its predictive accuracy and improve its usability. This final phase of research subjected the now finalised YORST (version 2) to a battery of standard psychometric tests important for such inventories. It also provided the opportunity to re-test the predicted accuracy of the Mini YORST,<sup>1</sup> an automatically generated risk score available to frontline police officers through their hand held mobility devices.

## Methods

The YORST(v2) was assessed for the following three psychometric properties.

- **Inter-rater reliability** – the degree of agreement among two or more individuals using the same instrument. If there is poor inter-rater agreement either the tool is defective or the raters need to be re-trained in the use of the tool. In this study, this was tested with multiple raters (72 Youth Aid Officers) reviewing a case study vignette designed by New Zealand Police.<sup>2</sup> A range of statistical tests were used to assess the degree of agreement in YORST risk scores among raters.
- **Concurrent validity** – this measures how well a particular test correlates with a previously established measure of the same construct (i.e. risk of re-offending). In the current study this was tested by 10 Youth Services Staff (Youth Aid Officers, Youth Development Officers or Child Case Managers) completing two risk assessments, (i) a YORST(v2) and (ii) a YLS/CMI 2.0<sup>3</sup> on one young offender. A total of 28 sets of risk assessments were completed using both tools and the overall level of association between the risk scores and similar sub-scales of the two scales was tested.
- **Predictive validity** – this is perhaps the most commonly discussed and arguably the most important type of validity in reference to risk screening/assessment tool. Predictive validity relates to a tool's ability to predict a criterion variable, which in this

---

1 Operationally this is referred to as the CYP Offending Risk Indicator.

2 To test inter-rater reliability, ideally you would replicate real life administration of YORST(v2) and use multiple raters on multiple cases. The approach selected was one that was more practical in terms of data collection and time demands required of Police Youth Services Staff.

3 The YLS/CMI 2.0 is one of the most widely used risk assessment tools for young offenders internationally with well-established psychometric properties.

research was a young offender's recidivism over 12 and 24 month periods. YORST(v2) scores were collected from a sample of 1212 young offenders who had completed a YORST(v2) between 1 March and 31 May 2013. The ability of the YORST(v2) total risk scores to accurately predict young offenders' recidivism over the following two years was tested along with the scores from the Mini YORST.

## **Inter-rater reliability and concurrent validity**

Tests of inter-rater reliability and concurrent validity of the YORST(v2) concluded the YORST(v2) had sound properties.

In regards to inter-rater reliability results, the YORST(v2) compared favourably to those presented in a recent study of the YLS/CMI 2.0 that used a similar methodology. Key findings included:

- all 72 raters scored the young person in the case study as 'medium risk' and 76% of the raters had total risk scores within 5 points (50 to 55 out of a possible 100). The average absolute deviation (ADD) was just 2.56 points away from the median total risk score
- reliability of individual items varied from near perfect agreement for two items (10 and 11), good agreement for three items (9, 13, & 8), to average for the final four (items 6, 14, 7 & 12). It could be these four items are more subjective, that raters require more training, or possibly the information detailed in the case study vignette was less clear in relation to these items.

In relation to concurrent validity, a perfect correlation between the YLS/CMI 2.0 and the YORST(v2) was not expected due to the differing content and emphasis in the two tools; however, overall results indicated a good level of association.

- total risk scores and appropriate sub-scales were all statistically significantly correlated. Total risk scores had a Pearson correlation coefficient of  $r=.552$ . This is considered a large effect size and has less than a 1 in 100 probability of occurring by chance ( $p<0.01$ ). Comparing sub-scales of the two tools, highest correlations occurred between the substance use items/sub-scales and those assessing education/employment
- both instruments assessed the majority of the young offenders as medium risk. However, according to the specific cut-off scores of each instrument for the various risk categories, the YORST categorised a greater proportion of the sample as low risk (21% compared to 14% with the YLS/CMI 2.0), whilst the YLS/CMI 2.0 had a greater dispersion of ratings.

## **Predictive ability of the YORST(v2) and Mini YORST**

The final, and most important validity test for the revised YORST(v2) was the ability of the total risk score to accurately predict recidivism in a sample of young offenders. The Mini YORST was subjected to a similar analysis.

A sample of 1212 YORST assessments completed between 1 March and 31 May 2013 were available for analysis. The majority of the sample were male (76%), aged 14 to 16 years old (73%) and were Māori (60%).

A range of indicators suggest that revisions made to the YORST(v2) have further enhanced the tool's predictive accuracy. The YORST(v2) produces accurate risk assessments that

now compare even more favourably to other well established more comprehensive youth risk assessment tools.

Key findings included:

- YORST(v2) accurately predicted 71% of those young offenders who were re-apprehended within 12 months, and 78% of those that were re-apprehended within 24 months
- there was a good linear relationship between YORST(v2) total risk scores and the percentage of those who were re-apprehended, with an increasing proportion of the sample re-apprehended as total risk scores increased. There was some evidence to suggest the creation of four risk categories may be a useful modification
- mean YORST(v2) risk scores were significantly higher for those young offenders who were re-apprehended within 12 and 24 months
- total risk scores had large, significant point bi-serial correlations with re-apprehension status within 12 months ( $r=.44$ ), and over 24 months ( $r=.42$ ), suggesting high levels of correlation between YORST total risk cores and recidivism
- survival analysis found that young offenders with higher YORST(v2) risk scores offended at a greater rate and sooner than those with lower scores, and indicated the tool does best at differentiating between low and medium risk offenders
- YORST(v2) significantly discriminated offenders who were re-apprehended within 12 and 24 months, with significant AUC scores of .759 and .774<sup>4</sup>
- the tool performed well across all demographic sub-groups including Māori offenders. This was an important improvement on the original YORST(v1)<sup>5</sup>
- higher YORST(v2) scores were also significantly associated with more frequent and more serious offending, and offending that occurred sooner
- the Mini YORST was also found to significantly discriminate offenders who were re-apprehended within 12 and 24 months, with significant AUC scores of .765 and .773. It also performed well across all demographic sub-groups.

## Item response analysis

A review of the performance of individual items found 7 out of the 14 items significantly contributing in predicting recidivism. Item 5 (total number of prior offences) had the greatest contribution in predicting 12 month recidivism, followed by item 2 (time since coming to Police notice for an incident) and item 7 (influential peers known to Police). Whilst other YORST(v2) items may appear redundant, it must be remembered that in addition to

---

4 These results are on par with those of other recognised tools. The AUC for the ASSET in predicting reconviction over 12 months was .712 (Baker et al., 2003) and .731 over 24 months. The YLS/CMI had a mean weighted AUC of .641 across 11 studies (Schwalbe, 2007) and the Australian version had an AUC of .65 over 12 months (Thompson and McGrath, 2012) and .75 over 24 months (Upperton and Thompson, 2007)

5 This is an important result as Māori young offenders are a key group for New Zealand Police to work effectively with. It is unlikely the modifications to the YORST could fully account for the improved performance of YORST with Māori. Whilst beyond the scope of this study to fully investigate, other explanations such as a change in police practice may also have had an influence.



predicting risk of recidivism, identifying areas of need that can be targeted through interventions is also of high importance.

Adding weighting to the significant items based on the logistic regression model was able to improve the accuracy of prediction (AUC = .781 compared to .759 for the more simple current cumulative risk assessment). The use of weighted scores may be worth considering in the future, however the potential gains in accuracy would need to be weighed up against high costs of revamping the current system of producing YORST scores and advantages of being able to score YORSTs manually using the current simpler cumulative risk assessment method..

The Mini YORST uses items 1, 2 and 5 in its calculation. Results from the current analysis suggest item 1 may now be redundant suggesting more analysis of the optimal items and weightings for the Mini YORST may be warranted.

Exploration of alternative predictor variables related to prior number and type of incidents confirmed their relationship with recidivism. However, analysis suggested the current YORST(v2) incident related item (item 2 – time since came to Police notice for an incident) was performing well and little would be gained by introducing more incident related variables.

## Conclusion

Reliable, evidence-based risk assessment tools are an essential component of work to ensure effective case management of the most at risk young offenders. Findings from this final phase of research have found on several psychometric properties considered important for such inventories, the YORST (v2) performed well and in keeping with other established inventories and this contributes to the confidence that can be placed in the integrity of the tool and its use by New Zealand Police.

Analysis demonstrated that the tool can be used reliably in New Zealand, with high levels of agreement among raters when scoring the same case. In terms of validity, the risk scores from the YORST(v2) were significantly associated with those from the YLS/CMI 2.0, which is perhaps the most widely used and validated risk assessment tool for young offenders.

Analysis of the tool's ability to predict recidivism found the revisions made to the YORST(v2) have further improved its predictive accuracy. The current tool now performs well for both male and female young offenders, those aged from 10 years through to 16 years, and for all major ethnic groups in New Zealand including European, Māori and Pacific.

In terms of future development, two areas are recommended for further investigation:

- according to current YORST(v2) cut-off points, the medium risk category is currently a very large group with correspondingly large variations in rates of recidivism. It would be sensible to explore if four risk categories are more useful than the current three (e.g., low, medium, high and very high)
- while the Mini YORST as currently designed is performing well, results from the current analysis suggest a review of the best items and weightings for the Mini YORST is warranted.



# 1 Introduction

This report presents the findings from the final phase of research to assess the validity, reliability and predictive capability of the New Zealand Police Youth Offending Risk Screening Tool (YORST). This was the third phase of a programme of research that was initiated in 2009. It included the testing of the revised YORST (version 2) and also the re-validation of the Mini YORST, an automatically generated risk score available to frontline police officers through their hand held mobility devices.

## 1.1 Background

Standardised risk and needs assessment and/or screening of offenders is well recognised overseas as best practice. New Zealand Police began work to develop a risk screening tool in 2003, resulting in the development and piloting of the YORST in 2007. The YORST was rolled out nationally in July 2009 following a favourable evaluation of the pilot.

In 2009 New Zealand Police secured funding through the Cross Departmental Research Pool to carry out a programme of research to assess the validity and reliability and predictive capability of the YORST. Dr Elaine Mossman who was then working for the Crime and Justice Research Centre (CJRC) at Victoria University of Wellington, was commissioned to carry out this research. This programme of research was divided into three phases:

- **Phase I** consisted of a literature review of international and New Zealand literature on risk assessment of young offenders and an analysis of the quality of existing YORST data. This work was completed in August 2010 (Mossman, 2010ab).<sup>6</sup>
- **Phase II** was an analysis of the predictive ability of the initial YORST (version 1) based on a sample of 1965 YORSTs completed between July and October 2009. This work was completed in November 2011 and resulted in recommendations on how the existing YORST could be modified/refined to improve its usability and predictive accuracy (Mossman, 2011). A revised YORST (version 2) was then developed and implemented in late 2012 (see Appendix 1).
- **Phase III** (the current report) has tested the validity of the revised YORST (version 2). This has included testing its psychometric properties (concurrent validity and inter-rater reliability) and the re-testing of the predictive accuracy following the modifications. It also provided the opportunity to re-test the predictive accuracy of the Mini YORST (an automatically generated risk score based on three of the YORST items, that is available to frontline police officers through their hand held mobility devices).<sup>7</sup>

---

6 Mossman, S.E. (2010a). Screening and assessment of young offenders risk of recidivism: Literature review. August 2010. Final report prepared for New Zealand Police, Wellington; Mossman, S.E. (2010b). Research to validate the YORST Phase I: Review of existing YORST data. August 2010. Final report prepared for New Zealand Police, Wellington.

7 Operationally the Mini YORST is referred to as the CYP Offending Risk Indicator.

## 1.2 Key findings from Phases I and II

### 1.2.1 Phase I (a) - Literature review

A comprehensive review of international and New Zealand literature summarised information from which to evaluate the attributes and content of the YORST including:

- factors found to be predictive of recidivism in young offenders
- important attributes of risk screening/assessment instruments
- methods for assessing the validity and reliability of instruments
- risk screening/assessment instruments used overseas and in New Zealand and their comparative effectiveness.

The review, in the main, supported the risk factors for recidivism included in the YORST although there were a few highly predictive risk factors identified in the literature that are not included (e.g. age of first placement in a youth justice residential facility, non-severe pathology, conduct problems, effective use of leisure time, length of first incarceration, number of prior commitments). Several of these were seen to be less relevant to the New Zealand youth justice system which emphasises rehabilitation over incarceration, and others may be considered less appropriate for police officers to assess (e.g. mental health status).

The inclusion of dynamic risk factors in addition to static risk factors was seen to be consistent with current best practice recommendations. However, it was noted that the YORST had been developed as a screening tool and as such was considerably shorter than other tools that aim to assess criminogenic need in addition to predicting risk of recidivism.

The review highlighted the importance of assessing the validity of tools such as the YORST for different sub-groups of young offenders (male and female, different ethnicities and ages).

### 1.2.2. Phase I (b) - Analysis of the quality of initial YORST data

This analysis provided:

- a profile of young offenders who were being screened using the YORST – which found the majority (70%) were classified as medium risk offenders
- a comprehensive assessment of missing data, including performance of automatically generated items – which found that missing data for individual items was variable across items, but not a major concern with 91% of YORSTs having fewer than two items missing. Some minor software problems were found with four of the automatically generated items that were subsequently addressed
- partial analysis of the accuracy and consistency of data being generated by the YORST in 2009 – analysis was limited by the unavailability of secondary sources to cross-check the accuracy of different YORST items. Where consistency and accuracy were able to be assessed, findings were mixed

- this work concluded with a review of the content and wording of each YORST item with recommendations for revision to be considered in conjunction with YORST phase II research findings.

### 1.2.3 Phase II – Predictive validity of the initial YORST (version 1)

Phase II carried out the following tasks with the final bullet being the primary focus:

- a usability assessment of the YORST (v1) to identify ways to improve quality of data collected by YORST
- a review of the primary purpose of YORST and its other objectives and how YORST information was being used
- **an assessment of the ability of existing YORST items and total YORST score to predict recidivism.**

A range of analyses concluded the ability of the initial YORST (version 1) to predict recidivism in young offenders compared favourably with other well established and typically more comprehensive risk assessment tools (e.g. UK's ASSET and YLC/CMI).

- YORST(v1) total risk scores correctly predicted re-apprehension status for 68.2% of the sample. The ASSET predicted 67% over a similar follow-up period.
- total YORST(v1) risk scores were significantly correlated with re-apprehension status ( $r=.34$ ), while other well established risk assessment tools had correlation coefficients ranging from .28 through to .43.
- overall accuracy of the YORST(v1) was found to be moderate to high (ROC analysis on the YORST (v1) produced an AUC of .695 for the entire sample and .703 for the Special Group of more reliable raters). Again the accuracy of the YORST (v1) compared well to other well established tools (for reconviction over 12 months, AUCs for the YLS/CMI ranged from .641 to .67, and .712 for the ASSET)
- the YORST(v1) performed equally well for males and females, but was more accurate for New Zealand European compared to Māori and also more accurate for older offenders and appeared better able to differentiate between low and medium risk offenders than it did between medium and high risk offenders.

Item-response analysis and logistic regression identified which items were most useful in predicting recidivism and outlined how modifications to the YORST(v1) through the re-coding of items, applying weights to predictive items or including additional predictor variables could improve its predictive ability.

These findings resulted in the minor revision of the YORST(v1) which included the re-coding of some items, and redundant items removed and substituted with alternative items that could improve the predictive ability of the YORST(v1). The automatic population of some items was also enabled (e.g. a young person's criminal history information). The new revised YORST(v2) went live in late 2012.

This phase II research also provided information that enabled the development and implementation of the **Mini YORST**, an automatically generated risk score available to frontline police officers through their hand held mobility devices. The Mini YORST risk score uses three automatically populated criminal history items from the full YORST, but with weights applied to increase its predictive power (weights identified through the phase II

logistic regression analysis). The three items selected were those that phase II research had identified as the best predictors of recidivism. The aim of the Mini YORST was to assist frontline officers to identify potentially high risk young offenders who could then be referred to specialist Youth Aid Officers for their cases to be reviewed and most appropriate action taken.

### 1.3 Objectives of Phase III

The revised YORST(v2) has now been in operation for over two years enabling the third and final phase III to be completed.

For a risk screening/assessment instrument to be able to improve the quality of youth justice decision-making it 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. Whilst **reliability** is the extent to which a test is consistently able to produce the same result across time and across raters.

Phase III assessed the YORST(v2) for the following three psychometric attributes (the predictive ability of the Mini YORST was also tested):

- **Inter-rater reliability** which is the degree of agreement among two or more individuals using the same instrument. If there is poor inter-rater agreement either the tool is defective or the raters need to be re-trained in the use of the tool.
- **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.
- **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 which in this case will be young offender's recidivism over 12 month and 24 month periods.

## 2 Methodology

### 2.1 Evaluation objectives

The overall aim of this research was to establish the psychometric properties of the revised YORST(v2) and re-validate the Mini YORST. The specific evaluation objectives of Phase III were to assess:

- if YORST(v2) has satisfactory **inter-rater reliability** when one or more individuals use the tool to assess the same young offender
- if the YORST(v2) has satisfactory **concurrent validity** when compared against another valid measure of risk
- to carry out item response analysis to review the ability of individual YORST(v2) items to predict recidivism and consider the value of an alternative predictor variable (total number of prior incidents)
- the accuracy of YORST(v2) total risk score in predicting recidivism over 12 month and 24 month periods
- the accuracy of the Mini YORST score in predicting recidivism over 12 month and 24 month periods on a sample of young offenders that was different from the one that was used to first develop it.

### 2.2 Details of evaluation methodology

The first component of phase III was to test the inter-rater reliability of the YORST(v2) and its concurrent validity.

#### 2.2.1 Concurrent validity and inter-rater reliability of finalized YORST(v2)

There were several methodological options available to test the concurrent validity and inter-rater reliability. Ideally you would replicate real life administration of YORST(v2) and use multiple raters on multiple cases for inter-rater reliability. However, such an approach is logistically challenging and resource intensive. The approaches selected and described below were more practical in terms of data collection and time demands required of Police Youth Services Staff and were possible to achieve with available allocated funds.

- **Inter-rater reliability** – was tested with multiple raters reviewing a case study vignette (this was similar approach used successfully by Lowenkamp et al, 2004).<sup>8</sup> The case study was designed by New Zealand Police and covered all areas represented on the YORST (see Appendix 2). The case study was emailed to 250 Youth Services staff around the country (aiming to collect 30-50 completed YORSTs from a sample of officers with varying levels of experience and geographic location), with each Youth Services staff then completing an online version of YORST based on the information presented in the case study. A total 72 surveys were completed

---

8 Lowenkamp, C, Holsinger, A, Brusman-Lovins, L & Latesa, E. (2004). Assessing the inter-rater agreement of the Level of Service Inventory Revised. *Federal Probation*, December 2004, 68(3), p 34-38.

and collated by New Zealand Police in consultation with the researcher. A range of statistical tests were used to assess the degree of agreement among raters.

- **Concurrent validity** - A research exercise was set up to assess concurrent validity. This consisted of 10 Youth Services Staff (Youth Aid Officers, Youth Development Officers or Child Case Managers) recruited to complete YORSTs on a young person, and at the same time also complete a YLS/CMI 2.0 inventory on the same young offender.<sup>9</sup> The aim was to complete 30-50 duplicate ratings. Ten youth services volunteers completed 28 sets of duplicate ratings on 28 young offenders. Comparing the association between results produced by the YORST(v2) and the YLS/CMI 2.0 tested concurrent validity. Pearson correlation tests were used to assess the overall level of association between the total risk scores and similar sub-scales of the two scales.

## 2.2.2 Predictive ability of finalised YORST(v2) and the Mini YORST

The final validity test for the revised YORST(v2) was the ability of the total risk score to accurately predict recidivism in a sample of young offenders. The Mini YORST was subjected to a similar analysis. The procedures and analyses largely replicated the validity testing carried out in Phase II on the initial YORST (v1) (Mossman, 2011) but this time on a new sample of young offenders and their respective YORST(v2) scores.<sup>10</sup>

### Sample

The sample of YORSTs for analysis was all those YORST(v2) collected in the three month period 1 March to 31 May 2013.<sup>11</sup> This resulted in a total of 1701 YORST(v2)s being retrieved. However, if a young person had more than one YORST completed in this period, just the first YORST was used for analysis.<sup>12</sup> YORSTs were also excluded from analysis if they were 'in progress' (n=42) or were Youth Development exit YORSTs or other Youth Development YORSTs (n=425) without an index offence / date of offending. This resulted in a final sample of n=1212. Mini YORST risk scores were computed from the YORST(v2) data.<sup>13</sup>

9 The YLS/CMI 2.0 was suggested as it requires just 40 minutes to complete and has been shown to produce valid results across different jurisdictions including Australia and New Zealand.

10 Mossman, S.E. (2011). *Predictive ability of the YORST*. Wellington: New Zealand Police. Accessed November 2011, <http://www.police.govt.nz/sites/default/files/yorst-predictive-ability-analysis.pdf>

11 The start date of 1 March 2013 has allowed for a two to three month bedding in period following the introduction on of YORST(v2) but also sufficient time for a 12 to 24 month follow-up period.

12 18 young offenders had more than one YORST completed accounting for 22 YORSTs results removed from analysis.

13 The 'B coefficients' from a logistic regression analysis were applied to the respective Mini YORST items together with the constant to calculate predicted probabilities of re-offending. The formula was  $P = e^{-2.105 + (\text{item1} \cdot 1.184) + (\text{item2} \cdot 1.159) + (\text{item5} \cdot 1.462)} / 1 + e^{-2.105 + (\text{item1} \cdot 1.184) + (\text{item2} \cdot 1.159) + (\text{item5} \cdot 1.462)}$ .



### Recidivism

The criminal history data (all NIA occurrence records) of the young offenders in the sample described above was extracted to assess recidivism in the 12 to 24 month period after the YORST(v2) was completed.<sup>14</sup>

The primary measure of recidivism was whether the young person has been **re-apprehended for a new offence** in the 12 month and 24 month periods.<sup>15</sup> A number of other recidivism measures were also assessed including:

- the frequency of re-offence
- the time to the first re-offence
- seriousness of re-offending.

The Lifetime Offenders Seriousness Tool (LOST) assisted in the identification and extraction of these recidivism variables.<sup>16</sup>

### Other considerations

1. **Historical offences** - the offence dates associated with new apprehensions were carefully reviewed, as new apprehensions for historical offending that occurred prior to the completion of the YORST had to be removed.
2. **Impact of youth justice interventions on recidivism** - if the YORST is administered prior to interventions in response to the offending, any positive impact of any intervention on recidivism will not be accounted for. It appears there is no easy way of obtaining information on exposure to interventions for this to be considered as a covariate. This is endemic to youth justice risk assessment in general but needs to be recognised as a limitation of this research also.
3. **Time at large** - it is important to ensure that each young offender has the same follow-up period in the community (time at large). The Children, Young Persons, and Their Families Act 1989 (the Act) provides for three main situations and associated orders where a young person can be held in custody:

---

14 Criminal history information prior to the YORST(v2) being completed was also used to examine further predictive variables (e.g. frequency of incidents).

15 The diversionary emphasis of the New Zealand Youth Justice system means a relatively small proportion of child (2%) and young (29%) offenders who come into contact with New Zealand Police are prosecuted. Therefore, while it is more common in young offender research to see reconviction (prosecution) data used to assess recidivism, re-apprehension data is more appropriate as it will provide recidivism data that is relevant to a greater proportion of youth that come to police notice. The other advantage of using re-apprehension data is that it is real time and is not subject to the procedural time delays of prosecution data where a Youth Court decision can take many months to be processed. The main limitation of this measure is that apprehensions represent the number of alleged offences and apprehensions do not necessarily mean that an offender has been charged or in the small number of cases where an offence is denied by a young person is then found to be not proven.

16 LOST is a Microsoft Excel based tool developed by New Zealand Police with assistance from the Ministry of Social Development that provides an automated process for summarising a person's offending history.

- detained in police custody following an arrest (section 236)
- remanded in custody pending a court hearing (section 238 1E)<sup>17</sup>
- sentenced to a supervision with residence order (section 311).

Assistance was sought from Child Youth and Family (CYF) to obtain details on any custody related court orders and the length of time of these orders for those young offenders in the sample. Where possible the number of days held in custody was added onto the 12 and 24 month follow-up periods to provide an equal 'time at large' (i.e. if a youth was in custody for 10 days, re-apprehension data was checked for 12 months and 10 days or 24 months and 10 days). CYF were provided with descriptive information on the young people of interest (name, age, location). This replicates methodology used in Phase II. Key analyses were run with and without taking into account time in custody.

4. **Retrieval of alternative YORST item data** - ideally the YORST(v2) should remain unchanged with its validation fully tested, however, based on requests by frontline officers the potential value of prior incidents was assessed.
5. **Confirmatory bias** - the YORST is administered at an early decision point within the criminal justice system with the aim of informing subsequent decisions and actions. However, there is a risk that results achieved by the YORST rather than informing, actually dictate subsequent actions (i.e. Police may decide to lay charges because of a high YORST risk score, which increases the criminal history of the young person and may result in increased criminal offending). This would mean the YORST scores become a "self-fulfilling prophecy" regardless of its validity rather than a reliable means to predict likelihood of recidivist behaviour. However, the predictive ability of the YORST(v2) is validated against new apprehension data for alleged offending which will be unrelated to the original YORST(v2) score being tested, although, it could be legitimately argued that subsequent apprehension was a result of increased attention given to high risk offenders who have had a YORST completed.

### Data cleaning and analyses

All data was incorporated into an excel spread sheet and exported to SPSS for analysis. Data was carefully screened for any inconsistencies or inaccuracies. Once cleaned a series of analyses were performed to assess the predictive ability of the YORST(v2) and Mini YORST:

- the percentage of cases of high risk offenders correctly classified<sup>18</sup>

---

17 Under section 238.1E a young person can be remanded in police custody for up to 10 days. Under section 238.1D a young person is usually remanded to a CYF secure residence, however, there are a few occasions where they may be remanded to a CYF community-based family home. Community-based family homes are very restrictive (e.g. curfews enforced) but not as secure as residences. It was difficult to determine which type of residence had been used and so for the purposes of the research custody resulting from 238.1Ds were not included.

18 YORST scores are split into high and low risk at a point corresponding to the proportions actually re-apprehended and all high scores are treated as predicting recidivism and low scores as predicting non-recidivism.

- the accuracy of the prediction across the range of risk scores (risk scores generated by the YORST plotted against the percentage of the sample re-apprehended). This also included the proportion correctly classified according to the current low, medium and high risk cut-off points.
- differences in mean risk scores of those who have been re-apprehended were compared to those who had not using an independent t-test
- survival analysis examined the ability of the YORST(v2) and Mini YORST to predict rates of recidivism and also gain more insight into time taken to be re-apprehended.
- finally the overall accuracy of the prediction was assessed using Area Under the Curve (AUC) Receiver Operating Characteristics (ROC) analysis. This analysis produces AUC scores ranging from 0.5 that indicate the tool is predicting no better than chance to 1.0 where the tool perfectly predicts recidivism.

It was important to assess the tool's ability to predict recidivism across different groups of young offenders (different ages, gender and ethnicity). This is particularly important considering the YORST (v1) was found to perform better with New Zealand European young offenders compared to Māori and was also more accurate for older offenders.

Item response analysis for phase III included:

- association of individual items with measures of recidivism to identify those most useful in predicting risk of recidivism
- logistic regression analysis to assess the relative contribution of different YORST items in predicting recidivism.<sup>19</sup> This analysis also provided an important validation check on the weighting currently applied to items used in the Mini YORST (it is important logistic model building is re-tested on a different sample to that used to establish initial weightings). It also enabled another review of the relative merit of using more sophisticated calculation of risk scores based on weighted items, compared to the current more simple cumulative risk model score.

---

<sup>19</sup> Note: some items may have roles other than purely predicting risk of recidivism e.g, identification of criminogenic needs to inform case management and interventions planning.

### 3 Inter-rater reliability and concurrent validity

For a risk screening/assessment such as the YORST(v2) to be able to improve the quality of youth justice decision-making it must be both valid and reliable. In this chapter these important psychometric properties of the YORST(v2) are examined.

#### 3.1 Inter-rater reliability

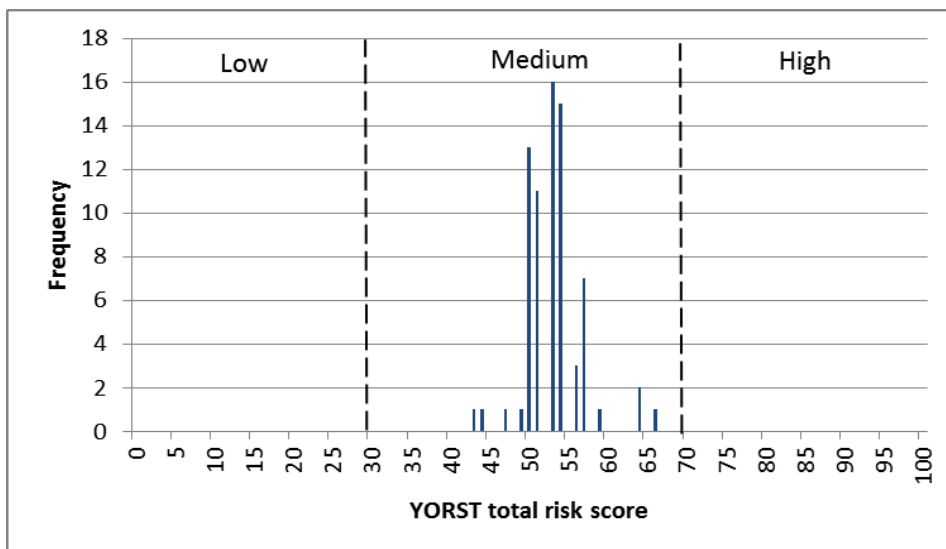
The inter-rater reliability of an instrument assesses the extent which a test is consistently able to produce the same result across raters. If there is poor inter-rater agreement either the tool is defective or the raters need to be re-trained in the use of the tool.

The inter-rater reliability of the YORST(v2) was tested by 72 volunteer raters scoring a YORST(v2) on the same case study vignette (see Appendix 2). Nearly all raters were Youth Aid Officers (n=69, 96%), with just three Youth Development staff. The majority had been acting in their role for more than 5 years (n=62, 86%), with a further eight in their role for between two to five years, and three under two years.<sup>20</sup> All 12 New Zealand Police Districts were represented.

##### 3.1.1 Inter-rater reliability of YORST(v2) Total risk score

The raters completed the nine subjective items on the YORST(v2), the remaining five of the criminal history factors are auto-populated and therefore did not require reliability testing. Figure 3.1 shows the distribution of total risk scores scored by the 72 raters. The YORST(v2) produces a total risk score from 0-100. To create meaningful total risk scores a fixed total of 13 was added to each rater's raw score to account for five auto-populated criminal history items before transforming scores into total risk scores (0-100).<sup>21</sup>

**Figure 3.1 Distribution of YORST total risk scores among 72 raters**



<sup>20</sup> The survey and case study was sent to all Youth Services staff (around n=250) of which 72 participated. There is little turnover amongst Youth Services staff and the level of experience of those who completed exercise is likely to be representative of Youth Services staff.

<sup>21</sup> 13 was the estimated score that would have been produced from items 1 to 5 based on the researcher's judgment of information in the case study.

Overall these initial results look good, all 72 raters had scored the young person in the case study as 'medium risk' and 76% of the raters had total risk scores within 5 points (50 to 55).

A statistical measure of inter-rater reliability sometimes used is the average absolute deviation (ADD). The ADD calculates the average variation around a measure of central tendency (in this case the median total score of 53 was used). The ADD was just 2.56 points away from median which again suggests reasonably good inter-rater reliability. This compares well to a recent inter-rater reliability test of the YLS/CMI where ADDs across 20 raters for 10 cases ranged from 1.04 to 4.95 and averaged 2.65 (Rocque & Plummer-Beale, 2014).<sup>22</sup> This is even more positive when you consider the ADD scores of the YORST(v2) total risk score would be expected to be larger due to the increased range of possible total scores 0-100 compared to 0-42 with the YLS/CMI.

Intra-class correlations (ICC) are a more rigorous and more commonly used measure of inter-rater reliability that enables agreement by chance to also be taken into account. Scores range from 0 to 1, with a score of 1 indicating perfect agreement. However, the current research had only one case study across multiple raters which meant the traditional use of this statistic was not possible. However, if the 72 ratings using each of the nine individual items are treated as separate assessments the test could be applied. When this was done the absolute agreement, single measures ICC was .836 ( $p < 0.001$ ), which suggests excellent reliability across items and raters. However, because this is not the traditional application of ICC results should be viewed with caution.

Taking these results together there appears to be reasonably good inter-rater reliability in relation to the YORST(v2) total risk score. Separate analyses were carried out based on length of service and role within Police. Numbers in some categories were small and again results need to be interpreted with caution, but were good for all groups. ICCs were slightly higher (more consistent) for Youth Development roles .954 compared to Youth Aid Officers .842; and slightly higher for those serving more than five years .851 compared to those serving five or less .814. ADDs showed a similar trend with smaller, more consistent results for Youth Development staff (1.6 compared with 2.6 for Youth Aid Officers) and those serving for over five years (2.5 compared with 3.0 for those serving five years or less).

#### **3.1.2 Inter-rater reliability of YORST(v2) individual items**

The performance of individual items can vary even when an overall score appears consistent and so it is important to consider the performance of the nine YORST(v2) items tested. Table 3.1 shows results for analysis of individual items, with those items with better reliability appearing first (low ADD and high percentage agreement). The mean and median statistics provide an indication of the level of risk as rated by the 72 raters, the ADD and percent agreement are the key measures in relation to inter-rater reliability. The number of response options is a variable that could influence the level of agreement. Having fewer options can assist in achieving agreement among raters.

---

22 Rocque, M & Plummer-Beale, J. (2014). In the eye of the beholder? An examination of the inter-rater reliability of the LSI-R and YLS/CMI in a correctional agency. *Journal of Criminal Justice* 42, 568-578.

**Table 3.1 Inter-rater reliability on individual YORST items**

Variable	Mean	Median	ADD	Percent agreement	No. of response options <sup>1</sup>
Item 10 – substance abuse	0.1	0	0.10	96%	4
Item 11 – evidence of family violence	3.04	3	0.13	94%	3
Item 9 – care and protection history	4.82	5	0.18	89%	5
Item 13 – parents criminal history	0.22	0	0.22	89%	5
Item 8 – education/employment	3.31	3	0.31	85%	5
Item 6 – concern re: (M/O) nature of crime	3.76	4	0.32	68%	5
Item 14 – sibling criminal history	2.49	3	0.51	63%	5
Item 7 – peers known to police	3.75	4	0.44	56%	5
Item 12 – other concerns re: family	2.71	3	0.68	53%	4

Table notes: <sup>1</sup> The number of response options can vary from 3 to 5 for different items, although the highest possible score is always 5.

There was variation in the reliability of ratings across the nine YORST items tested.

- **Item 10** and **item 11** have near perfect agreement.
  - Item 10 the extent to which current use of alcohol and/or other drugs contributes to anti-social/offending behaviour was judged as not present in 96% of cases (69 out of 72 raters).
  - Item 11 evidence of current and/or historical family violence achieved agreement in 68 out of 72 raters (94%). Of note this items has just three response options – no evidence (0), evidence of family violence in immediate family (3), and evidence of family violence convictions and/or court orders in immediate family (5).
- **Items 9, 13,** and **8** also had high levels of agreement (85-89%) and this is despite each having five response options.
- **Items 6, 14, 7** and **12** had lower levels of agreement and slightly higher ADD scores. It could be these items are more subjective, require more training, or possibly the information detailed in the case study vignette was less clear in relation to these items.
  - Item 6 is perhaps one where improved guidelines may help as to what constitutes concerning nature (M/O) of current offending
  - Item 14 appears clearly defined and was clearly described in the case study
  - Item 7 may benefit from clearer differentiation in the descriptors of number of influential peers known to Police e.g. ‘some’ compared to ‘many’ (although

analysis in chapter 5 suggests this item is currently performing quite well in predicting recidivism)

- Item 12, perhaps similarly to item 6, may benefit from clear guidelines on what constitutes 'other concerns' in regards to the family/living situation. Also perhaps whether concerns are current or historical, which may have been a confusing factor in the case study.

While not perfect, overall inter-rater results were more than satisfactory. Perhaps most importantly, all 72 raters placed the young person in the case study in the same 'medium' risk category. It is important however that limitations of the current test be recognised (i.e., this was not a real life test where raters would have to source data themselves to score the different items, and reliability was only tested for one type of offender).

## 3.2 Concurrent validity

Concurrent validity is a measure of how well a particular test correlates with a previously established measure of the same construct (i.e. risk of re-offending). In this case the YORST(v2) is tested against Hoge, Andrews and Leschied (2011) Youth Level of Service/Case Management Inventory 2.0 (YLS/CMI 2.0).<sup>23</sup> Good concurrent validity would be seen when risk ratings predicted by the YORST were associated with similar risk ratings on the YLS/CMI 2.0.

The YLS/CMI 2.0 is one of the most widely used risk assessment tools for young offenders with well-established psychometric properties (Hoge & Andrew, 2011).<sup>24</sup> Like the YORST(v2) it produces a total risk score and associated risk classification. Table 3.2 presents a breakdown of the content of the two tools and highlights that there are also some differences between the two tools. These include:

- the YLS/CMI 2.0 has four levels of risk classification (low, moderate, high, and very high). There are also different risk thresholds to estimate risk level for males and females. Unlike the YORST, the YLS/CMI 2.0 does not have a sub-total for dynamic risk
- as can be seen in table 3.2 the tools vary slightly in the sub-components assessed and their emphasis when added together to assess total risk. The YORST has a greater emphasis on offending characteristics of youth and their wider family/whānau (7 out of 14 items compared to 5 out of 42 in the YLS/CMI 2.0), whilst the YLS/CMI 2.0 has information collected on personality/behaviour, and constructive use of leisure time absent in the YORST. These differences mean a perfect correlation between total risk scores would not be expected
- the YLS/CMI 2.0 also has a greater number of individual items per sub-component, 42 in total compared with the 14 in the YORST. This is because the YORST has been designed as a screening tool as opposed to a full assessment tool and therefore you would expect fewer items.

---

23 Hoge, A., Andrews, D & Leschied, A. (2011). YLS/CMI 2.0 Youth Level of Service / Case Management Inventory. New York, United States: Multi-Health Systems.

24 Hoge, R & Andrews, D. (2011). YLS/CMI 2.0 Youth Level of Service/Case Management Inventory 2.0 User's Manual. Toronto, Canada: Multi-Health Systems.

**Table 3.2 Comparison of content of YORST and YLS/CMI 2.0**

Variable	YORST (14 items)	YLS/CMI 2.0 (42 items)
Offending history characteristics	✓ 5 items	✓ 5 items
Peer relations	✓ 1 item	✓ 4 items
Education/employment	✓ 1 item	✓ 7 items
Care and protection history	✓ 1 item	✗
Substance abuse	✓ 1 item	✓ 5 items
Family circumstances/parenting	✓ 2 items	✓ 6 items
Leisure/recreation	✗	✓ 3 items
Personality/behaviour	✗	✓ 7 items
Attitudes/orientation	(✓) <sup>1</sup> 1 item	✓ 5 items
Family criminal history	✓ 2 items	✗
Total risk score	✓	✓
Total dynamic risk score	✓	✗

Table notes: <sup>1</sup> In the YORST this relates to concern over the circumstances of the current offending.

### 3.2.1 Procedure

Ten New Zealand Police Youth Services staff were recruited for this study who were either Youth Development Officers or Child Case Managers. These two groups routinely conduct in-depth assessments with young offenders who are referred to them and as such it was expected they were more likely to already have sufficient information to complete the 42 items contained in the YLS/CMI 2.0. Raters were sent detailed instructions and scoring guidelines.<sup>25</sup> Each rater assessed between one and four young offenders with a total of 28 young offenders assessed using both the YORST and YLS/CMI. Of these, 15 were male and 13 female; 15 were Māori, ten New Zealand European and one was Pacific.<sup>26</sup> They ranged from 9 years to 15 years old.<sup>27</sup>

25 The scoring guidelines for 'prior and current offences/dispositions' of the YLS/CMI 2.0 were modified slightly to better relate to the diversionary emphasis of the NZ Youth Justice system. For the purposes of the research a 'conviction' was equated to any offence (not incident) recorded against an individual regardless of resolution (warned, diversion, intention to charge, court referred to youth justice family group conference, or any other court referral/resolution). This may have resulted in slightly higher total risk scores, but as long as all raters used the same definition would not impact on the concurrent validity exercise.

26 Two had no ethnicity specified.

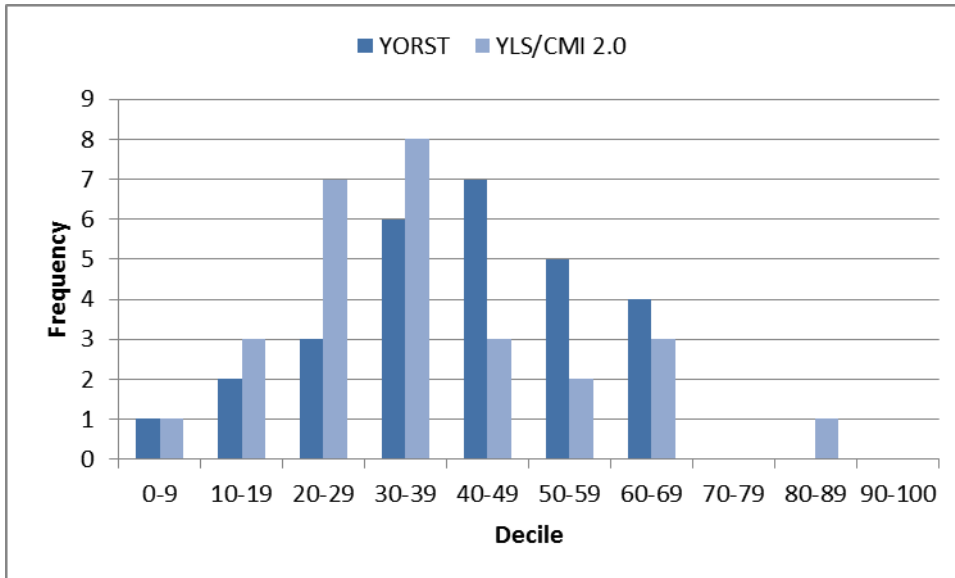
27 The YLS/CMI 2.00 is validated for youth 12 years and over, there were two youth in the sample who were under 12, however, results did not vary if these youth were excluded (correlation of total risk scores  $r=.552$ ,  $p<0.01$  if all ages included and  $r=.566$ ,  $p<0.01$  if excluded).



### 3.2.2 Total risk ratings

The distribution of the total risk scores from the two scales appears in figure 3.2 below (YLS/CMI 2.0 scores are usually reported as 0-42 but have been converted into percentages to allow for comparison).

**Figure 3.2 Distribution of total risk scores**



Note: YLS/CMI scores have been transformed and presented as percentages to be comparable to the YORST total risk scores (0-100).

The distributions of scores do vary in figure 3.2 with the YORST risk scores being slightly more negatively skewed and YLS/CMI 2.0 having a greater dispersion of ratings. However, according to the two instruments' specific cut-off scores for risk categories both instruments have the majority of the young offenders as medium risk.<sup>28</sup> The YORST had 21% of the sample as low risk (scores below 29) and 79% medium risk (scores 30-69). The YLS/CMI 2.0 had a greater dispersion of ratings, and whilst the majority were also medium risk (64%), fewer were low risk (14%), and more (21%) were high or very high risk.

As noted above, differences in the content of the two instruments will impact on how total risk is calculated. For example, higher scores are possible for the YLS/CMI 2.0 if there are personality/behavioural issues present or poor use of leisure time. In contrast if family members have a criminal history this could result in a higher YORST total score.

To help assess the degree of association between the two scales, YORST total risk scores (0-100) and YLS/CMI total risk scores (0-42) and where appropriate sub-components of respective tools were tested for their correlation. Results appear in table 3.3.

<sup>28</sup> The actual cut-off points for risk classification vary for the two instruments. Direct comparisons are further complicated by the YLS/CMI 2.0 having more risk categories and different cut-off points for males and females which had to be applied before comparisons could be made).

**Table 3.3 Correlation of YORST and YLS/CMI 2.0 (n=28)**

Variable	Correlation	Significance
<b>Total risk scores</b>	<b>.552</b>	<b>p&lt;0.01</b>
Offending history characteristics <sup>1</sup>	.513	p<0.01
Peer relations (YORST item 7)	.438	p<0.05
Education/employment (YORST item 8)	.615	p<0.001
Care and protection history	n/a	n/a
Substance abuse (YORST item 10)	.672	p<0.001
Family circumstances/parenting (YORST item 12) <sup>2</sup>	.596	p<0.01
Leisure/recreation	n/a	n/a
Personality/behaviour	n/a	n/a
Attitudes/orientation	n/a	n/a
Family criminal history	n/a	n/a

Table notes:

<sup>1</sup> The first six items of the YORST were combined to create a composite offending history score.

<sup>2</sup> If YORST items 9, 11 and 12 were combined into a composite family circumstances score, they had a similar correlation  $r=.555$ ,  $p<0.01$ .

Correlations between total risk scores and appropriate sub-scales were all statistically significantly correlated. Total risk scores had a Pearson correlation coefficient of  $r=.552$ , this is considered a large effect size and had less than a 1 in 100 probability of occurring by chance ( $p<0.01$ ). Highest correlations occurred between the substance use sub-scales and education/employment.

As noted earlier a perfect correlation was not expected due to the different content of the two scores, however, results above suggest the YORST has acceptable concurrent validity when tested against the YLS/CMI 2.0.

### 3.3 Summary

In summary, results presented in this chapter suggest the YORST(v2) has reasonably sound psychometric properties.

In regards to inter-rater reliability results for the YORST(v2) compared favourably to those presented in a recent study of the YLS/CMI 2.0 that used a similar methodology.

Key findings included:

- all 72 raters scored the young person in the case study as 'medium risk' and 76% of the raters had total risk scores within 5 points (50 to 55). The average absolute deviation (ADD) was just 2.56 points away from the median total risk score
- reliability of individual items varied from near perfect agreement for two items (10 and 11), good agreement for three items (9, 13, & 8), to average for the final four (items 6, 14, 7 & 12). It could be these four items are more subjective, raters require more

training, or possibly the information detailed in the case study vignette was less clear in relation to these items.

In relation to concurrent validity, a perfect correlation between the YLS/CMI 2.0 and the YORST(v2) was not expected due to the differing content and emphasis in the two tools, however, overall results were largely positive.

- both instruments had the majority of the young offenders as medium risk. However, according to the specific cut-off scores of each instrument for the various risk categories, the YORST categorised more of the sample as low risk (21% of scores below 29) and medium risk (79% of scores 30-69). In comparison, the YLS/CMI 2.0 had a greater dispersion of ratings, and whilst the majority were also medium risk (64%), fewer were categorised as low risk (14%), and more (21%) were high or very high risk
- correlations between total risk scores and appropriate sub-scales were all statistically significantly correlated. Total risk scores had a pearson correlation coefficient of  $r=.552$ , this is considered a large effect size and had less than a 1 in 100 probability of occurring by chance ( $p<0.01$ ). Highest correlations occurred between the substance use sub-scales and education/employment.

## 4 Predictive ability

The first section of this chapter describes the demographic characteristics and recidivism behaviour of those young offenders who had a YORST(v2) screen carried out between 1 March and 31 May 2013. This provides important descriptive information on the sample of youth on whom subsequent analysis is based. Following this there are a series of analyses that test the accuracy of the YORST(v2) and the Mini YORST to predict the recidivism behaviour of this group. The current research was able to test the YORST(v2) against 24 month recidivism in addition to 12 month recidivism which had been carried out previously when testing the original YORST(v1).

### 4.1 Sample characteristics

As described in the methods section (chapter two) a sample of 1212 YORST assessments were available for analysis. Table 4.1 presents the demographics of the youth who were the subject of these YORST(v2)s. The right hand side of the table reports what percentage of each demographic sub-group recidivated in the 12 months and 24 months after completed their YORST(v2). The overall rate of recidivism for the sample in the first 12 months following the YORST (V2) was 64.4% (n=781) of the sample and this increased to 75.4% (n=914) over 24 months. The 12 month recidivism rate was similar to that of the first YORST(v1) validation research (66.3% re-apprehended).

**Table 4.1 Demographics of research sample**

Variable		Demographics of sample		Recidivism of sample (re-apprehension)	
		n	%	Percent in 12m (n=781)	Percent in 24m (n=914)
<u>Gender</u>	Male	923	76%	68%	79%
	Female	289	24%	54%	64%
<u>Age</u>	Under 10 years	18	2%	6%	22%
	10 to 13 years	311	26%	61%	72%
	14 to 16 years	880	73%	67%	78%
	Over 16 years	3	0.2%	33%	67%
<u>Ethnicity</u>	Māori	727	60%	68%	77%
	NZ European	306	25%	62%	76%
	Pacific	138	11%	62%	77%
	Asian	9	1%	33%	44%
	Other	10	1%	80%	90%
	Unknown	22	2%	14%	18%
<b>Total</b>		<b>1212</b>	<b>100%</b>	<b>64%</b>	<b>75%</b>

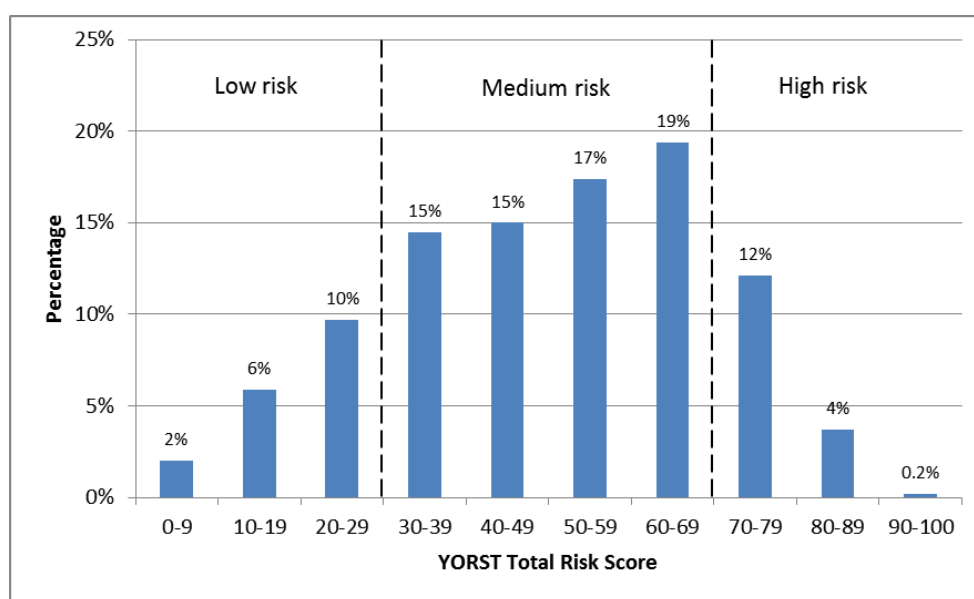
Table notes: Percentages are rounded and may not add up to 100%. Figures in grey are based on low frequencies and are not reliable.

From the left hand side of the table you can see the majority of the sample were male (76%), aged 14 to 16 years old (73%) and were Māori (60%).

Looking at the rates of recidivism for different demographic sub-groups it can be seen over both 12 months and 24 months that males had higher rates of recidivism than females, and older youth (14 to 16 years) had slightly higher rates than those in younger groups. Māori had slightly higher rates of recidivism over 12 months, but had similar rates to other groups over 24 months.

Figure 4.1 presents the distribution of YORST(v2) scores for this sample. There is a slight negative skew towards medium to high risk. The YORST(v2) categorised the majority of the sample as medium risk (66%), with 18% considered low risk and 16% high risk.

**Figure 4.1 YORST(v2) total risk scores of sample (n=1212)**



## 4.2 Percentage of offenders correctly predicted

To calculate the percentage of cases correctly predicted by the YORST(v2) the same methodology described by Baker et al., 2003 in the testing of the UK's risk assessment tool the ASSET was utilised. This involved splitting YORST(v2) scores into high and low risk at a point corresponding to the proportions actually re-apprehended (64.4%). For predicting 12 months recidivism, the highest 64.4% of YORST(v2) scores were considered high risk scores and assumed to predict recidivism (i.e. re-apprehension); whilst the lowest 35.6% of YORST total risk scores were assumed to predict non-recidivism. High scorers re-apprehended, and low scores not re-apprehended are counted as correctly predicted, and the rest as incorrect. A similar method was used for assessing 24 months accuracy. Results appear in tables 4.2 and 4.3 and overall numbers of cases correctly predicted were as follows:

- 12 months follow-up – the YORST(v2) accurately predicted 71.4% of cases correctly (21.3% plus 50.1%)
- 24 months follow-up – the YORST(v2) accurately predicted 77.6% of cases correctly (13.4% plus 50.1%).

**Table 4.2 Percentage of cases correctly predicted at 12 months**

	Not re-apprehended 12 months		Re-apprehended 12 months	
	n	%	n	%
Low risk	258	21.3%	173	14.3%
High risk	173	14.3%	608	50.1%
Total	431	35.6	781	64.4

Note: Bolded figures in shaded cells represent those cases predicted correctly (i.e. 21.3% plus 50.1% = 71.4% correct).

**Table 4.3 Percentage of cases correctly predicted at 24 months**

	Not re-apprehended 24 months		Re-apprehended 24 months	
	n	%	n	%
Low risk	162	13.4%	136	11.2%
High risk	136	11.2%	778	64.2%
Total	298	24.5%	914	75.4%

Note: Bolded figures in shaded cells represent those cases predicted correctly (i.e. 13.4% plus 64.2% = 77.6% correct).

These results are a slight improvement over the original YORST that correctly predicted 68.2% of cases and again compares well to the UK's ASSET which is far a more comprehensive 106 item full risk assessment tool. Over 12 months the ASSET correctly predicted reconviction for 67% of a group of n=1081 young offenders (Baker et al., 2003). After 24 months, the accuracy of the ASSET to predict reconviction increased to 69.4% (Baker et al., 2005).

### 4.3 Accuracy of the prediction across the range of total risk scores

Total YORST risk scores can range from 0 to 100. The original developers of the YORST categorised scores into low, medium or high risk based on the following cut-off points:

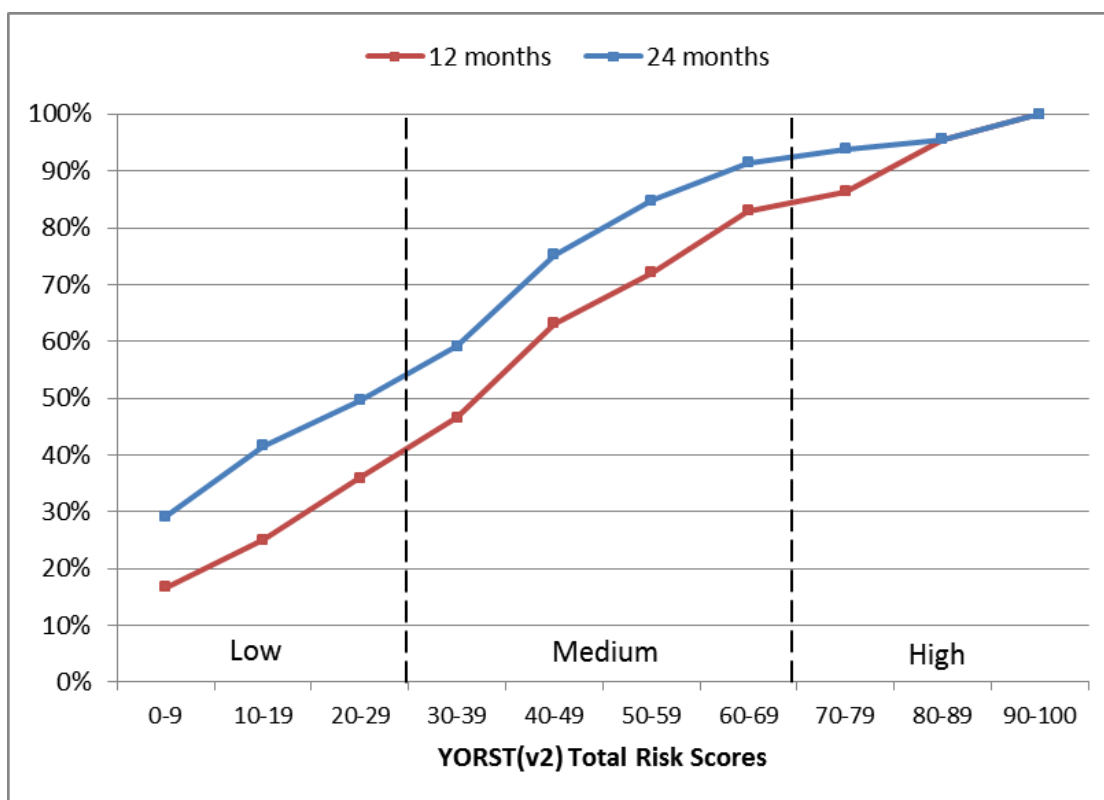
- Low risk = total YORST risk scores of 0 to 29
- Medium risk = total YORST risk scores of 30 to 69
- High risk = total YORST risk scores of 70 to 100.

Table 4.4 presents the percentage of cases re-apprehended for these three risk categories. The current cut-off points appear to do reasonably well at differentiating between low, medium and high risk groups.

**Table 4.4 Percentage of cases re-apprehended across low, medium and high YORST(v2) total risk scores**

	Re-apprehended 12 months		Re-apprehended 24 months	
	n	%	n	%
Low risk (n=213)	64	30%	95	45%
Medium risk (n=804)	544	68%	635	79%
High risk (n=195)	173	89%	184	94%

The accuracy of the tool was further assessed by comparing the percentage of the sample re-apprehended at 12 months and 24 months for each of ten YORST(v2) risk score bands going up in ten point groupings. Results appear in Figure 4.2.

**Figure 4.2 YORST(v2) scores against percentage re-apprehended**

The graph shows a steady increase in the proportion of the sample re-apprehended at 12 months as YORST(v2) total risk scores increased. Again this is an improvement from the original YORST. Looking at rates of re-apprehension over 24 months, after a steady increase, scores of over 60 show a slight flattening off. This suggests when looking at longer term recidivism, the cut-off point for high risk may currently be on the high side with little difference in recidivism of those with scores of 60-69 (medium risk) and those with scores 70-79 (high risk). Looking back at 12 months recidivism there are a high percentage of those with scores 60-69 who are re-apprehended (83%) and it could be

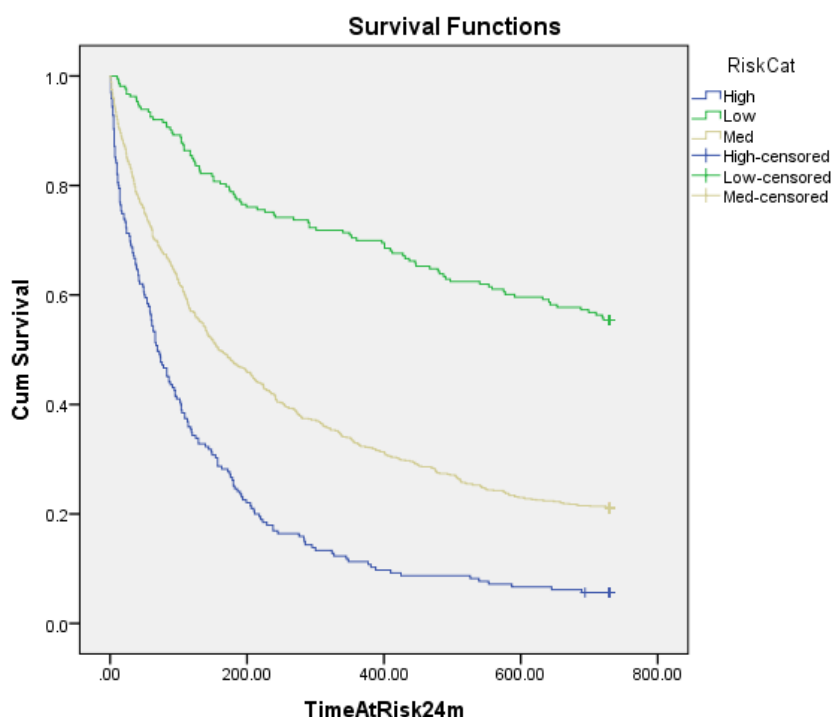
worth considering creating an extra category similar to the YLS/CMI 2.0 (low, medium, high, very high).

## 4.4 Survival analysis

To further examine the ability of YORST(v2) to predict recidivism and to gain insight into time taken to recidivate, survival analysis was carried out on rate of re-apprehension over 24 months as a function of 'time at risk'. This analysis plots the proportion of offenders that have not been re-apprehended (survived) as a function of time at risk, and permits all cases to be used for the time they are at large.<sup>29</sup>

Figure 4.3 presents the Kaplan-Meier survival function for recidivism using the current YORST cut-off points for low (0-29), medium (30-69) and high risk (70-100) scores (as described in the previous section).

**Figure 4.3 Kaplan-Meier Survival Function for 24 month recidivism for young offenders classified as low, medium or high risk according to YORST(v2)**



A log rank test was used to compare survival time (i.e. time without a re-apprehension) of the three groups (low, medium and high risk) and the difference was highly significant,  $\chi^2 = 190.4$ ,  $p < 0.0001$ , suggesting the YORST(v2) scores significantly differentiate time to first offence as well as likelihood of recidivating.

This plot clearly shows that as a group offenders classified as high risk by YORST(v2) (the lower line) are more likely to offend and offend sooner than those with lower total risk scores (the top line). Whilst statistically different, the plots representing the medium and high risk groups are closer together than the medium and low risk groups. This

<sup>29</sup> For non-recidivists time at risk was calculated as 730 days minus any days spent in secure custody, for recidivists it was the time until first re-apprehension.



provides some evidence that the current YORST(v2) cut-off points are better able to differentiate between low and medium risk offenders, with differences between medium and high risk, evident but less marked. Mean survival times were 145, 286 and 524 days for high, medium and low risk groups respectively.

#### 4.5 Mean YORST(v2) scores of those re-apprehended compared to those not

Another way to check the accuracy of the YORST(v2) was to compare the difference in mean total YORST risk scores of those who had been re-apprehended, to those that had not. Table 4.5 presents the results of an independent t-test showing there was a highly significant difference in total YORST risk scores between the two groups ( $p < 0.001$ ).

**Table 4.5 Difference in mean total YORST risk scores of those re-apprehended and those not re-apprehended**

	Mean (sd)	t	df	p
Re-apprehended (12m)	55.7 (17.0)	16.8	1210	<.001
Not re-apprehended (12m)	38.3 (17.8)			
Re-apprehended (24m)	54.1 (17.6)	15.9	1210	<.001
Not re-apprehended (24m)	35.5 (17.1)			

#### 4.6 Point bi-serial correlation

When comparing the ability of a tool to predict recidivism, the point bi-serial correlation is another common statistic.<sup>30</sup> The correlation of YORST(v2) total risk score against recidivism within 12 months was  $r = .44$  and a similar  $r = .42$  against recidivism over 24 months. These are higher than the correlation achieved for the initial YORST ( $r = .34$ ) over 12 months, and now compare even more favourably to other well established more comprehensive risk assessment tools. Olver et al., 2009 carried out a meta-analysis of three tools: Youth-adapted Psychopathy Checklist (PCL), the Structured Assessment of Violence Risk in Youth (SARVY) and the Youth Level of Service Case Management Inventory (YLS/CMI). The mean weighted  $r$  for predicting general recidivism for these tools was  $r = .33$  for the SARVY (across 7 studies),  $r = .32$  for the YLS/CMI (across 19 studies) and  $r = .33$  for the PCL (across 20 studies). The Australian version of the YLS/CMI-AA had an  $r = .26$  for predicting reconviction over 12 months (Thompson and McGrath, 2012) and  $r = .43$  for 24 months reconviction (Upperton and Thompson, 2007).

30 Point bi-serial correlation is the same as an ordinary correlation except one of the variables is binary with only two possible outcomes.

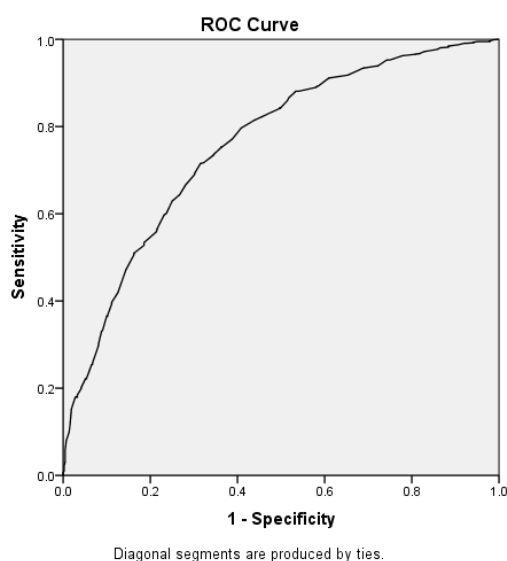
## 4.7 Receiver Operating Characteristics (ROC) analysis

In addition to point bi-serial correlation the most commonly used technique to assess and compare the accuracy of risk assessment/screening tools is to carry out Receiver Operating Characteristics (ROC) analysis. A ROC graph plots the proportion of young people in the sample correctly identified as recidivists (sensitivity) against the proportion of offenders incorrectly identified as recidivist (1-specificity), across a range of different cut-off risk scores used to classify offenders as recidivists.

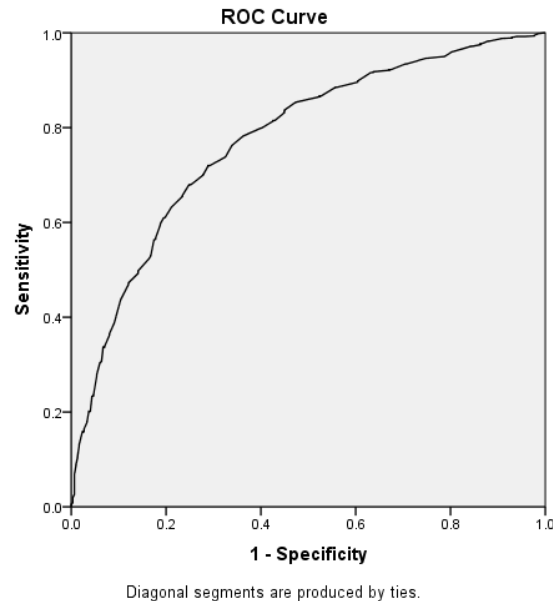
The area under the ROC curve (AUC) is an index of a tool's overall accuracy with scores produced ranging from 0 to 1 (0.5 indicates a chance-level accuracy; below 0.5 indicates a below-chance accuracy with tools incorrectly classifying the majority of offenders; above 0.5 indicates accuracy above-chance). A score of 1.0 indicates perfect discriminant accuracy or no false positive error. AUCs for an acceptable screening tool have been reported as being between 0.70 and 0.90 (Swets, 1988: cited in Vincent et al. (2009). Rice and Harris (2005) have produced tables to compare AUCs to other common measures of effect size (or significance of a statistical finding). They report that an AUC of .639 is equivalent to a moderate effect size, and one of .714 or higher is equivalent to a large effect size (highly significant).

ROC analysis was carried out on the YORST(v2) total risk scores to assess how accurately these were able to predict whether a young person was re-apprehended in 12 and 24 months following their risk assessment. Results reported are based on analysis of equivalent time at large. ROC analyses were also run without taking time at large into account, but results were unchanged. Figure 4.4 and 4.5 shows the ROC curves obtained. The AUC statistic for 12 months re-apprehension was .759, a large effect size according to Rice and Harris (2005). This can be interpreted as a 75.9% probability that a randomly selected recidivist would score higher on the YORST(v2) total score than a randomly selected non-recidivist. This is a good improvement on the original medium effect score of .659 with the original YORST. The results for YORST(v2) improved further in predicting re-apprehension over 24 months, with AUC being .774.

**Figure 4.4 ROC curve for the YORST(v2) total risk score over 12 months (n=1212); AUC=.759**



**Figure 4.5 ROC curve for the YORST(v2) total risk score over 24 months (n=1212); AUC=.774**



Predicting recidivism in young offenders is not an easy task (shorter histories to draw on than adults) and none of the tools do exceptionally well at this. However, the revised YORST(v2) continues to compare well to other risk assessment tools. The AUC for the ASSET in predicting reconviction over 12 months was .712 (Baker et al., 2003) and .731 over 24 months. The YLS/CMI had a mean weighted AUC of .641 across 11 studies (Schwalbe, 2007) and the Australian version had an AUC of .652 over 12 months (Thompson and McGrath, 2012) and .75 over 24 months (Upperton and Thompson, 2007).

The YORST(v2) also performed well in predicting medium seriousness and violent offending over both 12 and 24 months but no better than any offending. There was no improvement in results if prediction was limited to offences considered of medium seriousness or higher (see table 4.6).<sup>31</sup>

Finally ROC analyses were carried to see how well the tool performed for young offenders with different demographic characteristics. Results appear in the lower half of table 4.6.

31 The same methodology as the first study was used to determine medium serious offending, offences with seriousness scores higher than 67.8 were considered medium or higher. Seriousness scores for each offence were calculated using the Ministry of Justice 2012 offence seriousness scoring.

**Table 4.6 ROC analyses for YORST (v2) total risk scores<sup>32</sup>**

	12 months		24 months	
	AUC	95% CI	AUC	95% CI
<b>Complete sample (n=1212)</b>				
Any offending	.759	(.730, .787)	.774	(.744, .804)
Violent offending	.698	(.664, .732)	.710	(.679, .740)
Medium seriousness offending	.744	(.716, .771)	.758	(.731, .785)
<b>Demographic sub-groups</b>				
Males (n=923)	.769	(.736, .801)	.790	(.754, .826)
Females (n=289)	.727	(.669, .785)	.731	(.671, .790)
Māori (n=727)	.738	(.700, .777)	.747	(.705, .789)
New Zealand European (n=306)	.775	(.699, .811)	.786	(.731, .840)
Pacific Island (n=138)	.805	(.729, .881)	.841	(.758, .925)
10 to 13 years (n=311)	.747	(.690, .804)	.755	(.692, .817)
14 to 16 years (n=880)	.761	(.728, .795)	.778	(.742, .814)

Table notes: All AUCs were statistically significant ( $p < 0.0001$ )

AUC statistics suggest the existing YORST appears to predict recidivism well for all demographic sub-groups across 12 and 24 months. All scores would be considered large effect sizes and were statistically significant. This is an improvement on the original YORST that performed considerably less well for Māori (AUC=.661) and younger offenders 10 to 13 years (AUC=.679). This is an important result as Māori young offenders are a key group for New Zealand Police to work effectively with. It is unlikely the modifications to the YORST could fully account for the improved performance with Māori. Whilst beyond the scope of this study to fully investigate, other explanations such as a change in police practice may also have had an influence.

## 4.8 Relationship between YORST and other measures of recidivism

Further analysis found higher YORST scores were also statistically significantly associated with the frequency, severity and offending that occurred sooner:

- more frequent offending ( $r = .377$ ,  $p < 0.001$ )
- more serious offending ( $r = .260$ ,  $p < 0.001$ )
- offending that occurred sooner ( $r = -.468$ ,  $p < 0.001$ ).

<sup>32</sup> ROC analyses were also run without taking time at large into account. All key results were unchanged.

These results were obtained by correlating YORST(v2) total risk scores with the total number of offences over the next 12 months, the single most serious offence using the Ministry of Justice seriousness scale, and the number of days to the first recorded offence. Similar results were achieved against offending that occurred over 24 months.

## 4.9 Accuracy of the Mini YORST

This validation research of the revised YORST(v2) also represented an opportunity to re-validate the performance of the Mini YORST on a new sample of youth against their recidivism over 12 and 24 months.

Mini YORST scores were calculated as per current practice using YORST items 1, 2 and 5. Predicted probability scores are calculated with Beta coefficients from an earlier logistic regression analysis applied to the respective Mini YORST items together with a constant. The formula used to calculate Mini YORST scores is:

$$P = e^{-2.105 + (\text{item } 1 \cdot .184) + (\text{item } 2 \cdot .159) + (\text{item } 5 \cdot .462)} / 1 + e^{-2.105 + (\text{item } 1 \cdot .184) + (\text{item } 2 \cdot .159) + (\text{item } 5 \cdot .462)}$$

The AUC statistic for 12 months re-apprehension was .765, slightly higher than the full YORST (.759) and than the initial AUC score when the Mini YORST was first developed (.736). The AUC statistic for 24 months was a slightly higher AUC=.773. Encouragingly as can be seen in table 4.7, using this new sample the Mini YORST performed well for the different demographic sub-groups including Māori. All demographic sub-groups had large effect sizes.<sup>33</sup>

While the Mini YORST is performing very well, it is not performing significantly better than the full YORST which does not have weights applied. It could be the items and weighting used in the Mini YORST could be revised to further enhance its predictive accuracy. There will be further discussion of this in the next chapter.

---

33 Rice and Harris (2005) have produced tables to compare AUCs to other common measures of effect size (or significance of a statistical finding). They report that an AUC of .639 is equivalent to a moderate effect size, and one of .714 or higher is equivalent to a large effect size (highly significant).

**Table 4.7 ROC analyses for Mini YORST**

	12 months		24 months	
	AUC	95% CI	AUC	95% CI
<b>Complete sample (n=1212)</b>				
Any offending	<b>.765</b>	<b>(.738, .793)</b>	<b>.773</b>	<b>(.743, .802)</b>
Violent offending	.681	(.646, .716)	.682	(.651, .714)
Medium seriousness offending	.734	(.706, .762)	.744	(.716, .771)
<b>Demographic sub-groups</b>				
Males (n=923)	.772	(.740, .804)	.785	(.750, .821)
Females (n=289)	.773	(.675, .790)	.726	(.667, .784)
Māori (n=727)	.749	(.711, .786)	.755	(.715, .796)
New Zealand European (n=306)	.756	(.703, .810)	.770	(.715, .825)
Pacific Island (n=138)	.791	(.712, .870)	.809	(.711, .907)
10 to 13 years (n=311)	.759	(.705, .813)	.757	(.667, .816)
14 to 16 years (n=880)	.766	(.734, .799)	.774	(.738, .809)

Table notes: All AUCs were statistically significant ( $p < 0.0001$ )

## 4.10 Summary

A range of indicators suggest that revisions made to the YORST(v2) have further enhanced the tool's predictive accuracy. Predicting recidivism in young offenders is not an easy task (shorter histories to draw on than adults) and none of the tools do exceptionally well at this. However, the revised YORST(v2) now compares even more favourably to other well established more comprehensive youth risk assessment tools. Key findings included:

- YORST(v2) accurately predicted 71% of those young offenders who were re-apprehended within 12 months, and 78% of those that were re-apprehended within 24 months
- there was a good linear relationship between YORST(v2) total risk scores and the percentage of those who were re-apprehended. There was some evidence to suggest the creation of four risk categories may be a useful modification
- mean YORST(v2) scores of young offenders who were re-apprehended within both 12 and 24 months were significantly higher than those that did not
- total risk scores had large, significant point bi-serial correlations with re-apprehension status within 12 months ( $r = .44$ ), and over 24 months ( $r = .42$ )

- survival analysis found that young offenders with higher YORST(v2) risk scores offended at a greater rate and sooner than those with lower scores
- YORST(v2) significantly discriminated offenders who were re-apprehended within 12 and 24 months, with significant AUC scores of .759 and .774. The tool performed well across all demographic sub-groups including Māori offenders. These results look good in comparison to other tools. The AUC for the ASSET in predicting reconviction over 12 months was .712 (Baker et al., 2003) and .731 over 24 months. The YLS/CMI had a mean weighted AUC of .641 across 11 studies (Schwalbe, 2007) and the Australian version had an AUC of .67 over 12 months (Thompson and Pope, 2005) and .75 over 24 months (Upperton and Thompson, 2007)
- higher YORST(v2) scores were also significantly associated with more frequent and more serious offending, and offending that occurred sooner
- the Mini YORST was also found to significantly discriminate offenders who were re-apprehended within 12 and 24 months, with significant AUC scores of .765 and .773. It also performed well across all demographic sub-groups.

## 5 Item response analysis

There are 14 items in the YORST(v2) that are added together to make up the total YORST risk score, three of which are included in calculation for the Mini YORST risk score. While Chapter four assessed the accuracy of the total risk scores of the YORST(v2) and Mini YORST, this chapter reviews the performance of the individual YORST(v2) items in predicting recidivism and considers if additional information on prior incidents could be useful alternative predictors.

### 5.1 Item analysis – frequency of responses

The distribution of responses (or endorsement proportions) to individual items can impact on their usefulness in predicting recidivism. For instance, if over 90% of the sample are all receiving the same rating for an item (e.g. all have peers that are known repeat offenders – item 7) this item would have limited scope in being able to differentiate between individuals of higher or lower levels of risk.

Previous analysis on the YORST items found good distributions. Results from the current analysis remained good. Overall proportions for any response rating were all well below 90%; they ranged from 0.8% (item 8 – young person in irregular employment) through to 60.1% (item 11 – evidence of family violence in immediate family).

The level of missing data, (where items had not been completed) had improved since the previous report. In the current sample of the 1212 young offenders, there were 1081 (89%) with no missing data, and just 13 (1%) with three or more items missing. Item 14 (sibling criminal history) was the item most frequently not entered, with 5% of cases missing (n=64 out of 1212).

### 5.2 Inter-correlation between items

The next analysis assessed the degree to which individual items were correlated with recidivism and also with each other. Those with high correlations with re-apprehension are likely to be the most useful in predicting recidivism. Those with high levels of inter-correlation indicate potential redundancy among items. Table 5.1 provides the correlation of YORST(v2) items 1 to 14 with each other, and also with the two outcome measures re-apprehension for any offence in 12 month and the 24 month follow-up periods (correlations with outcome measures are in bold).

As with previous analysis all items were significantly and positively correlated with recidivism ( $p < 0.01$ ), with often slightly stronger correlations with 12 months than 24 months re-apprehension. The statistical significance of these results are partly due to the large sample size and it is important to consider the actual strength of the correlation coefficient ( $r$ ). Those shaded have a correlation coefficient of greater than 0.25 (considered a moderate correlation). The number of shaded boxes shows a high level of inter-correlation between items suggesting likely redundancy among items in terms of their unique contribution to the prediction of recidivism. This is particular true for the static prior offending items (e.g. items 1 and 5; and items 3 and 5).



**Table 5.1 Correlations with YORST(v2) items 1 to 14 with each other and recidivism (12m and 24m)**

	12mRcd	24mRcd	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14
12mRcd	1	.769**	.372**	.342**	.325**	.192**	.409**	.215**	.324**	.259**	.264**	.216**	.107**	.196**	.121**	.149**
24mRcd	-	1	.326**	.323**	.325**	.209**	.399**	.215**	.287**	.227**	.236**	.213**	.140**	.199**	.123**	.140**
Item 1	-	-	1	.422**	.487**	.178**	.805**	.400**	.482**	.314**	.283**	.399**	.128**	.316**	.130**	.197**
Item 2	-	-	-	1	.290**	.101**	.422**	.285**	.345**	.297**	.359**	.326**	.244**	.356**	.179**	.171**
Item 3	-	-	-	-	1	.305**	.688**	.317**	.416**	.255**	.286**	.346**	.124**	.240**	.106**	.185**
Item 4	-	-	-	-	-	1	.338**	.133**	.196**	.081**	.281**	0.008	.199**	.184**	.160**	.238**
Item 5	-	-	-	-	-	-	1	.417**	.496**	.325**	.326**	.397**	.164**	.315**	.158**	.222**
Item 6	-	-	-	-	-	-	-	1	.419**	.345**	.329**	.410**	.140**	.441**	.156**	.161**
Item 7	-	-	-	-	-	-	-	-	1	.384**	.315**	.456**	.180**	.428**	.211**	.293**
Item8	-	-	-	-	-	-	-	-	-	1	.289**	.390**	.212**	.370**	.182**	.136**
Item 9	-	-	-	-	-	-	-	-	-	-	1	.254**	.361**	.471**	.260**	.265**
Item 10	-	-	-	-	-	-	-	-	-	-	-	1	.116**	.405**	.149**	.125**
Item 11	-	-	-	-	-	-	-	-	-	-	-	-	1	.375**	.431**	.261**
Item 12	-	-	-	-	-	-	-	-	-	-	-	-	-	1	.317**	.320**
Item 13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	.288**
Item 14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1

Table notes: A double asterisk (\*\*) indicates correlations are statistically significant at  $p < .01$ . Those shaded have the strongest correlation with  $r > .25$ . Items 1, 2, 3, 5, 7, 8 and 9 have strong correlations of  $r > .25$  with both 12 month and 24 month recidivism.

The strongest individual predictors in order of their strength were item 5 – count of prior offences, item 1- time since came to notice for offending, item 2 – time since last came to notice for an incident, item 3 – highest level of previous intervention, and item 7- antisocial peers. These items were also highly correlated with each other, along with many of the other items.

With highly inter-related items it is useful to identify their unique contribution once the relationship explained by other items has been accounted for (i.e. eliminate the redundancy among items). The logistic regression analysis which follows assesses this.

### 5.3 Logistical regression

The YORST(v2) items 1 to 14 were entered into a forward stepwise (wald) logistic regression analysis to explore which were the best predictors of whether a young person was re-apprehended within 12 months. Results appear in table 5.2.

**Table 5.2 Logistic Regression analysis of YORST(v2) items**

Variables in the equation	B (SE)	Wald	df	Sig	Exp(B)
Constant	-1.873 (.240)	60.746	1	.000	.154
Item 2 (time since last incident)	0.211 (.046)	21.191	1	.000	1.235
Item 4 (age of first offence)	0.138 (.06)	5.316	1	.021	1.148
Item 5 (no. of prior offences)	0.271 (.045)	35.705	1	.000	1.311
Item 7 (peers)	0.144 (.051)	7.907	1	.005	1.155
Item 8 (education)	0.128 (.049)	6.831	1	.009	1.136
Item 9 (care and protection)	0.124 (.044)	7.774	1	.005	1.132
Item 11 (family violence)	-0.106 (.049)	4.576	1	.032	0.9

Table notes: % correctly predicted = 74.6;  $R^2 = .295$  (Nagelkerke) .214 (Cox & Snell); Model Chi-square=260.59,  $p < .001$ .<sup>34</sup>

In comparison to the original YORST(v1), more of the YORST(v2) items (7 out of 14) are now seen to be uniquely useful in predicting recidivism (only four were significant with the original YORST(v1)). For this set of results, the odds ratios (Exp[B]) provide the simplest way to assess the relative contribution of each item. An odds ratio of greater than 1 indicates that as the predictor increases, the odds of the outcome occurring increases. Conversely, a value of less than 1 indicates that as the predictor increases, the odds of the outcome occurring decreases (i.e. is a protective factor rather than a risk factor). Item 5 had the greatest contribution in predicting re-apprehension with the highest odds ratio of 1.311, this was followed by item 2 (odds ratio =1.235) and then item 7 (odds ratio=1.155). Interestingly whilst item 11 on its own is positively correlated with recidivism (see table 5.1), in the current model its odds ratio was below 1 and therefore, its additional contribution is as a protective factor.

<sup>34</sup> Logistic regression does not have an equivalent to the R-squared that is found in multiple regression; instead pseudo-R-square statistics are calculated Nagelkerke  $R^2$  and Cox & Snell  $R^2$ . These statistics do not mean what R-squared means in multiple regression (the proportion of variance explained by the predictors), but still provide a gauge of the substantive significance of the model, with larger values indicating a better fitting model.

Results suggest for this sample of young offenders, the automatically generated items are still the most useful in predicting recidivism particularly item 5 (number of prior offences) and item 2 (time since last came to police notice for an incident). However, compared with previous results, the unique contribution from item 1 appears now to be redundant perhaps explained by all useful variance now being accounted for by item 5.<sup>35</sup> The subjectively rated dynamic factors that were significant included the number of influential peers known to police as before but now also item 8 – engagement in educational/employment. Whilst other items may appear redundant, it must be remembered that in addition to predicting risk of recidivism, identifying areas of need that can be targeted through interventions is also of high importance.

The same analysis was run against 24 months re-apprehension and produced similar results. Models were very similar,  $R^2 = .303$  (Nagelkerke)  $.203$  (Cox & Snell); Model Chi-square=245.22,  $p < .001$ , with 80.1% correctly predicted. Item 5 and item 2 remained the most influential, item 11 and item 9 were no longer significant, and item 3 (highest level of previous intervention) became significant.

### **Weighting of items based on logistical regression model**

SPSS can use the logistic regression model reported in table 5.2 to calculate the probability of an event occurring (i.e. re-apprehension within 12 months). These estimated probabilities for each case are equivalent to risk scores. A ROC analysis of these predicted probability scores allows assessment of whether a more sophisticated calculation of risk scores using different weightings of items based on the logistic regression model is superior to the current more simple cumulative risk assessment where responses to items are added together. The predicted probabilities using this model did improve the AUC results (.781 compared to .759). However, using weighted scores for calculating YORST risk scores would be a major change to the current method that can be completed without the assistance of a computer. However, this may be worth revisiting in the future.

## **5.4 Mini YORST weightings**

The previous chapter found the total risk score produced by the Mini YORST was performing well in predicting both 12 and 24 months recidivism. However, the Mini YORST uses items 1, 2 and 5 in its calculation. Results from the current analysis suggest item 1 may now be redundant. More analysis of the best items and weightings for the Mini YORST may be warranted.

## **5.5 Alternative predictor items**

Frontline police officers had queried whether the total number and type of prior incidents might be an additional useful predictor. In their experience, they felt they were often aware of young person coming to police notice for other incidents (e.g. truancy, missing person, family violence) before later coming to attention for their own offending. Certainly the significance of item 2 (time since came to notice for an incident) was evident in the logistic regression analysis.

---

35 Item 1 and item 5 are highly correlated but do not breach the logistic regression assumption of multicollinearity (VIFs under 5.0 and tolerance values over .10).

## 5 Item response analysis

Four possible incident-related variables were considered as alternative, all were calculated based on the total number of incidents in the previous five years.

- i. total number of prior incidents (1D, 1G, 1H, 1J, 1K, 1T, 1X, 2M, 2W, 6C, 6D, 6E)
- ii. total number of prior family violence incidents (1Ds),
- iii. total number of failure to comply incidents – where there had already been offending (e.g. 2W-warrants, and 6D (bail breach) and 6E (electronic monitoring bail breach))
- iv. total number of prior incidents excluding the failure to comply incidents.

As a first step the individual correlations of these variables with 12 and 24 month recidivism were examined (see bolded results).

**Table 5.3 Correlations of incident variables (total number) with recidivism (12m and 24m)**

	12mRcd	24mRcd	(i) All prior	(ii) FV(1Ds) only	(iii) Compliance only	(iv) All except (iii)
12mRcd	1	<b>.769**</b>	<b>.250**</b>	<b>.123**</b>	<b>.164**</b>	<b>.214**</b>
24mRcd	-	1	<b>.220**</b>	<b>.133**</b>	<b>.149**</b>	<b>.185**</b>
(i) All prior	-	-	1	<b>.416**</b>	<b>.656**</b>	<b>.853**</b>
(ii) FV(1Ds) only	-	-	-	1	<b>.124**</b>	<b>.458**</b>
(iii) Compliance only	-	-	-	-	1	<b>.166**</b>
(iv) All except (iii)	-	-	-	-	-	1

Table notes: A double asterisk (\*\*) indicates correlations are statistically significant at  $p < .01$ . Those shaded have the strongest correlation with  $r > .25$ .

All variables are significantly correlated with recidivism, with (i) total number of prior incidents in the previous five years having the strongest correlation ( $r = .250$ ).

The first three variables were entered into a forward stepwise (wald) logistical regression along with the YORST(v2) items to explore if they could usefully add anything to the prediction of re-apprehension over 12 months. Item (iv) was not included as it was a derivative of (i) and/or (iii) and hence breached the logistic regression assumption of multicollinearity. Results are presented in table 5.4.

**Table 5.4 Logistic Regression analysis of YORST(v2) items and possible alternative predictors**

Variables in the equation	B (SE)	Wald	df	Sig	Exp(B)
Constant	-1.762 (.244)	60.746	1	.000	0.172
Item 2 (time since last incident)	0.173 (.049)	12.402	1	.000	1.188
Item 4 (age of first offence)	0.133 (.06)	4.929	1	.026	1.142
Item 5 (no. of prior offences)	0.255 (.046)	31.039	1	.000	1.291
Item 7 (peers)	0.141 (.051)	7.601	1	.006	1.152
Item 8 (education)	0.113 (.049)	5.198	1	.023	1.119
Item 9 (care and protection)	0.103 (.045)	5.129	1	.024	1.108
Item 11 (family violence)	-0.11 (.049)	5.000	1	.025	0.896
Variable (i) No. prior incidents	0.04 (.019)	4.334	1	.037	1.041

Table notes: % correctly predicted = 74.9; R<sup>2</sup> = .300 (Nagelkerke) .218 (Cox & Snell); Model Chi-square=265.681, p=.000.

Only one of the alternative predictors was able to significantly add anything extra to the current YORST(v2) items, this was the total number of prior incidents in the previous five years. However, it was the weakest of the eight selected items, and the overall model was only marginally better than presented in table 5.2 (Nagelkerke R<sup>2</sup> = .300 compared to .295 without the additional variable). It is likely most of the variance explained by the new alternative incident variables is already being accounted for with item 2.

## 5.6 Summary

Item response analysis found a good distribution of responses for the 14 items across the sample of young offenders and all items were significantly and positively correlated with 12 and 24 month recidivism ( $p < 0.01$ ).

Logistical regression analysis found Item 5 (total number of prior offences) had the greatest contribution in predicting 12 month recidivism, followed by item 2 (time since coming to Police notice for an incident) and item 7 (influential peers known to Police). Whilst other YORST(v2) items may appear redundant, it must be remembered that in addition to predicting risk of recidivism, identifying areas of need that can be targeted through interventions is also of high importance.

A more sophisticated calculation of risk scores using different weightings of items based on the logistic regression model was able to improve the accuracy of prediction (AUC = .781 compared to .759 for the more simple current cumulative risk assessment). The use of weighted scores may be worth considering in the future, however the potential gains in accuracy would need to be weighted up against high costs of revamping the current system of producing YORST scores and advantages of being able to score YORSTs manually using the current simpler cumulative risk assessment method.

The Mini YORST uses items 1, 2 and 5 in its calculation. Results from the current analysis suggest item 1 may now be redundant suggesting more analysis of the best items and weightings for the Mini YORST may be warranted.

## 5 Item response analysis

---

Alternative predictor variables related to prior number and type of incidents were examined to explore if their addition could further improve the accuracy in predicting recidivism. Whilst these variables were associated with later recidivism, their inclusion added little more than was already being explained by the current YORST(v2) incident related item (item 2 – time since came to Police notice for an incident).

## 6 Conclusion

Reliable, evidence-based risk assessment tools are an essential component of work to ensure effective case management of the most at risk young offenders. Findings from this final phase of research have found on several psychometric properties considered important for such inventories, the YORST (v2) performed well and in keeping with other established inventories and this contributes to the confidence that can be placed in the integrity of the tool and its use by New Zealand Police.

Analysis demonstrated that the tool can be used reliably in New Zealand, with high levels of agreement among raters when scoring the same case. In terms of validity, the risk scores from the YORST(v2) were significantly associated with those from the YLS/CMI 2.0, which is perhaps the most widely used risk assessment tool for young offenders.

Analysis of the tool's ability to predict recidivism found the revisions made to the YORST(v2) have further improved its predictive accuracy. Encouragingly the current tool was found to perform well for both male and female young offenders, those aged from 10 years through to 16 years, and for all major ethnic groups New Zealand European, Māori and Pacific Island.

In terms of future development there are two areas that warrant consideration:

- according to current YORST(v2) cut-off points, the medium risk category is currently a very large group with correspondingly large variations in rates of recidivism. It would be good to explore if four risk categories are more useful than the current three (low, medium, high and very high)
- the Mini YORST as it is currently calculated is performing well, however, results from the current analysis suggest a review of the best items and weightings for the Mini YORST may be warranted.

## 7 References

- 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.
- Hoge, A., Andrews, D & Leschied, A. (2011). YLS/CMI 2.0 Youth Level of Service / Case Management Inventory. New York, United States: Multi-Health Systems.
- Hoge, R & Andrews, D. (2011). YLS/CMI 2.0 Youth Level of Service/Case Management Inventory 2.0 User's Manual. Toronto, Canada: Multi-Health Systems.
- Lowenkamp, C, Holsinger, A, Brusman-Lovins, L & Latesa, E. (2004). Assessing the inter-rater agreement of the Level of Service Inventory Revised. *Federal Probation*, December 2004, 68(3), p 34-38.
- Mossman, S.E. (2010a). Screening and assessment of young offenders risk of recidivism: Literature review. August 2010. Final report prepared for New Zealand Police, Wellington.
- Mossman, S.E. (2010b). Research to validate the YORST Phase I: Review of existing YORST data. August 2010. Final report prepared for New Zealand Police, Wellington.
- Mossman, S.E. (2011). *Predictive ability of the YORST*. Wellington: New Zealand Police. Accessed November 2011, <http://www.police.govt.nz/sites/default/files/yorst-predictive-ability-analysis.pdf>
- 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.
- Rice, M. , and Harris, G. (2005). Comparing Effect Sizes in Follow-up Studies: ROC Area, Cohen's d, and r. *Law and Human Behavior*, 29 (5), 615-620.
- Rocque, M & Plummer-Beale, J. (2014). In the eye of the beholder? An examination of the inter-rater reliability of the LSI-R and YLS/CMI in a correctional agency. *Journal of Criminal Justice* 42, 568-578.
- Schwalbe, C. S. (2007). 'Risk assessment for juvenile justice: A meta-analysis'. *Law and Human Behavior*, 31 (5): 449-462.
- Thompson, A., P. & McGrath, A. (2012). Subgroup differences and implications for contemporary risk-need assessment with juvenile offenders. *Law and Human Behavior*, 36(4), 345-355.
- 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.
- 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.
- 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.



# Appendix 1: YORST (v2)

<b>Youth Offending Risk Offending Tool</b>				<b>V2</b>	<b>YORST ID</b>	12345678910
V2-20120611A						
<b>NAME</b>	Young		Person		<b>NIA Person ID Master PRN</b>	12345678910
<b>(Child/YP)</b>	Surname		First name(s)		<b>DOCLOC#</b>	1234/1234
<b>DOB</b>	23/03/2001	<b>Age</b>	14	<b>Gender</b>	Male	
<b>ETHNICITY</b>	European	Pacific	Asian	Other		
				<b>Date RST Completed</b>		<b>Status</b>
				10/08/2015		Completed
<b>Maori</b>	Iwi		Hapu			
<b>Incident / Offence Code</b>			<b>Incident / Offence Description</b>			
<b>YORST Type</b>	Standard			<b>Date of Offending</b>	To	
<b>CYF Site Office</b>				<b>Proposed Youth Aid Response</b>		

Part (A) Offending Factors					
Time since <b>last</b> came to Police notice for their <b>offending</b>					
No Previous	Over 2 Years	Over 12 Months and up to 2 Years	Over 6 Months and up to 12 Months	Over 1 Month and up to 6 Months	Less than 1 Month
0	1	2	3	4	5
Time since last came to Police notice for <b>incidents/tasks</b> relating to them					
No Previous	Over 2 Years	Over 12 Months and up to 2 Years	Over 6 Months and up to 12 Months	Over 1 Month and up to 6 Months	Less than 1 Month
0	1	2	3	4	5
Highest level of previous intervention (final outcome)					
No Previous	Noting	Warning	Alt. Action	FGC	Court
0	1	2	3	4	5
Age of first offence (recorded by Police)					
No Offences	Over 14 Years Old	14 Years Old	13 Years Old	10 To 12 Years Old	Under 10 Years Old
0	1	2	3	4	5
Count of prior offences					
No prior offences		1 to 3 prior offences		4 or more prior offences	
0		3		5	
Concern about nature (MO) of current offending?					
Very Low	Low	Medium	High	Extreme	
1	2	3	4	5	
<b>Comments re Question 6</b>	Young person was recently the perpetrator of an assault. This matter is now with YA and will be dealt with accordingly. There has been improvements in some areas of his offending - however refuses to address his aggressive behaviour.				

## Appendices

### Part (B) Peer Group Factors

Influential peers known to Police

None	Very Few Known	Some Known	Many Known	All Known Repeat Offenders	Unknown
0	1	3	4	5	Not scored

Comments re Question 7

This young person continues to socialise with anti social peers, who are known to Police.

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

Youth is under 16

#### 8a Current engagement in Education/Training

Full time - well engaged	Full time - some issues	Irregular attendance and/or behaviour problems	Stood down / suspended	Non-Enrolled / Excluded
0	1	3	4	5

#### 8b Current engagement in Employment

Full time - well engaged	Part time	Irregular employment	Unemployed (seeking employment)	Unemployed - not seeking employment
0	1	2	3	4

Enrolment Type	Enrolled in School	Effective Date	10/08/2015
----------------	--------------------	----------------	------------

Reason non-enrolled	
---------------------	--

Tertiary Institution / Course	
-------------------------------	--

School			Alternative Education Programme
--------	--	--	---------------------------------

Employment Status		Employment Start Date	
-------------------	--	-----------------------	--

Comments re Question 8

This young person is currently enrolled in Alternative Education and maintains good attendance. He was out of the education system for approximately 12 months and has settled back in well.

**Part (D) Care & Protection History**

Has a care and protection notification been made to CYF for this family or child / young person?

No	C&P notification concerning another sibling made by anyone	Notification concerning this young person made by anyone	Some form of C&P intervention provided by CYF for this young person	Currently or previously in C&P custody of CYF - 101 Status
0	2	3	4	5

Comments re Question 9

Young person is currently 101 with CYF

**Part (E) Alcohol and/or Drug Use**

10 Extent to which current use of alcohol and/or other drugs contributes to anti-social/offending behaviour?

Not at all	Slightly	Moderately	Significantly
0	1	3	5

Comments re Question 10

Young person reported that alcohol and drugs are apart of his lifestyle, however he has significantly reduced his intake.

**Part (F) Family Factors**

11 Evidence of current and/or historical family violence?

Nothing	Evidence of family violence in immediate family	Evidence of family violence convictions and/or court orders in immediate family
0	3	5

12 Other concerns about family/living situation?

None	Low	Medium	High
0	2	3	5

Detail Concerns

History of neglect and domestic disputes. Concerns for living situation are on-going - remains 101 status

13 Parental involvement in crime (current = within the last 12 months, historical = greater than 12 months)

None	Historical offending	Current offending	Historical imprisonment	Current imprisonment
0	2	3	4	5

14 Sibling involvement in crime (current = within the last 12 months, historical = greater than 12 months)

None	Historical offending	Current offending	Current youth court orders	Historical imprisonment	Current imprisonment
0	2	3	4	4	5

Any General Comments

Young person's mother has been subject to criminal activity within the last 12 month period, including common assault, other threatening behaviour.

Appendices

Information Sources					
Spoken to	Child / Young Person	Parent / Caregiver	School / Course Provider / MOE	Child Youth & Family	Other Agency
		This Time	This Time	This Time	This Time

Scoring Instructions						YORST Score
Questions			Answers			
Questions	No.	Max. Score	Current Risk Screening Total	=	43	x 100 = <b>61%</b>
Answered	14	x 5	Max. Total for Answered Questions	=	70	
Unanswered	0	x 5	Max. Total for Unanswered Questions	=	0	

Dynamic / Static Risk Factors					Dynamic YORST Score
Static Factor Results		No.	Static Factor Results		
Total for Static Factors		22	Total for Dynamic Factors		21
Max Possible Total for Static Factors		35	Max Possible Total for Dynamic Factors		35

Creating and Last Modifying Details		
Date Created	Time Created	Creating User
10/08/2015	10:11:24	
Last Modified Date	Last Modified Time	Last Modifying User
10/08/2015	14:14:28	

## Appendix 2: Case study vignette

### YORST SCORING CASE STUDY

**John Stacey DOB 20/12/98 (age 16)**

#### Circumstances of Offences

- John has multiple driving and unlawful taking of motor vehicle charges committed over an eleven month period from May 2014 until April 2015. John has also been associated with a number of recent burglaries.
- John explains his offending as being motivated by an intense desire to drive vehicles, and enjoys drifting and driving at high speed in the local industrial area. John says that he can strip down, repair, and rebuild almost any car engine.
- John stated that he was always with older friends (a number of whom have significant offending history) when he took cars. John takes responsibility for his actions but says he is not into burglary, which was always his friends' idea. They would ask him if he would drive and he would say yes to keep in with them.
- John has reflected on his action over the last year and is deeply remorseful. Meeting a victim at the Family Group Conference made John realise that his actions have caused great distress. He also realises the hurt he has inflicted on his girlfriend, and on his mother, Suzy, both whom are strongly against offending.

#### Personal History of Young Person

- John is the child of Richard and Mere; however he has been raised by Richard's sister Suzy after she lost her only child. John has lived with Suzy all his life and regards her as his mother.
- John suffered a traumatic sexual attack when he was eight years old and soon after his behaviour deteriorated markedly at school. Suzy struggled to cope and inflicted excessive physical discipline on him.
- Custody of John was granted in favour of the Chief Executive of the Department of Child Youth and Family Services on 28 September 2012. This followed from evidence that Suzy was unable to suitably parent John due to her own personal difficulties and failure to accept intervention to help her and John.
- John spent 18 months living with his uncle Bill while still in custody of the Chief Executive.
- Following substantial lifestyle changes made by Suzy, and the desire by John to return to his mother's care, they were reunited in 2014 and CY&F ended its custody order.
- Suzy has recently said that she is very upset with her son's behaviour as prior to last year he was attending church with her ever week and seldom in trouble.
- Suzy added that once John started college he started associating with his older cousins and their friends (some of whom have an offending history) and the trouble started. She said John knows right from wrong but has been influenced by his older associates who encourage him to drive.

- John currently lives with Suzy and his half sister Allannah.
- John states that he is very close to his mother Suzy, and his younger half sister Allannah.
- Allannah has come to Police attention twice in the last 12 months for shoplifting from the local dairy. Suzy has no offending history.
- John is also close to his father, Richard, who John sees regularly. John also sees his Uncle Bill as a strong support.
- John said that his girlfriend Nikita is a positive person in his life as she is still at college, hard working and does not break the law.
- There have been a number of 1D incidents involving Suzy and her boyfriend Phil, who occasionally stays at Suzy's house.
- John has a long history of defiance and disobedience at school, dating back to his traumatic time growing up around abuse and being victim of a sexual attack.
- John last attended school at Smith's College in 2014. Following his increased truancy he was eventually referred to the Ministry of Education's NETS division. He was referred to an alternative education provider in September 2014, but indications are that he only attends sporadically.
- John says that he is very keen to obtain employment and/or enter training - his goal is to be a diesel mechanic.
- John has a learners drivers licence, although is currently disqualified.
- John enjoyed boxing at school and would like to pursue this further if there is a local club. John has also developed an interest in Rugby League, and he likes to go to the gym.
- His main passion is with driving cars, and he is hopeful that a legal 'drifting' race track will be open to the public in Tauranga soon.
- John has a strong aversion to alcohol and drugs. He said he has never even had a cigarette. John said he has never tried any drugs but he has tried beer but did not like it. His mother she said that John had probably been put off by watching her misuse in the past.
- John said that his friends largely respect his stand on alcohol and drug use.