It is the paper published as:

Abstract: The predictive validity for the Australian Adaptation of the Youth Level of Service/Case Management Inventory was tested in a large sample (N = 4,401) of community-based juvenile offenders in New South Wales, Australia. First, we compared gender and ethnic subgroups on domain, total scores, and predictive validity. Both similarities and modest differences emerged in mean scores across subgroups. The pattern of predictive validity results showed comparable indices by gender and ethnic subgroups. Second, we supplemented our quantitative method with a review of 26 case files with the lowest risk scores and a 1-year reoffense, and 25 case files with the highest risk scores and no 1-year reoffense. We discuss implications of the findings for improving the predictive validity and practical utility of risk–need assessment with juvenile offenders.

Differentiating predictive validity and practical utility for the Australian Adaptation of the Youth Level of Service/Case Management Inventory (YLS/CMI-AA)

Andrew J. McGrath, Anthony P. Thompson, and Jane Goodman-Delahunty

Charles Sturt University

Author note

Andrew McGrath, School of Psychology, Charles Sturt University; Anthony P. Thompson, Jane Goodman-Delahunty, School of Psychology, Charles Sturt University.

Correspondence concerning this article should be addressed to Andrew McGrath, School of Psychology, Charles Sturt University, Building C6, Panorama Avenue, Bathurst, NSW 2795, Australia. Email: amcgrath@csu.edu.au

Abstract

The predictive validity for the Australian Adaptation of the Youth Level of

Service/Case Management Inventory (YLS/CMI-AA) was tested in a large sample (n

= 4401) of community-based juvenile offenders in New South Wales (NSW),

Australia. In the first part of our study we compared gender and ethnic subgroups on

YLS domain and total scores, and on predictive validity. Both similarities and modest

differences emerged in mean scores across subgroups. The pattern of predictive

validity results showed comparable indices by gender and the same was largely the

case for ethnic subgroups. In the second part of this research, we supplemented our

quantitative method with a review of case files for 26 of the lowest risk scoring cases

with a 1-year reoffense and for 25 of the highest risk scoring cases with no 1 year

reoffending. We discuss the qualitative methodology and draw attention to the

implications of the findings for improving the predictive validity and practical utility

of risk-need assessment with juvenile offenders.

Key words: Juvenile Offending, Risk Assessment, Recidivism, Youth Level of

Service/Case Management Inventory

Differentiating predictive validity and practical utility for the Australian

Adaptation of the Youth Level of Service/Case Management Inventory (YLS/CMI-

AA)

The structured assessment of risk and needs has been an active area of interest within

juvenile and adult corrections over the past 30 years. Risk-need assessment is

considered best practice within the field and the overarching risk-need-responsivity

framework (Andrews et al., 1990) is a correctional framework that is widely accepted

and practiced (Andrews, Bonta, & Wormith, 2006; Andrews et al., 1990). Numerous

risk-need measures exist but the Level of Service/Case Management Inventory for

adults (LS/CMI; Andrews, Bonta, & Wormith, 2004) and the Youth Level of

Service/Case Management Inventory (YLS/CMI; Hoge & Andrews, 2011) are well

reputed and widely used. Over one million officially declared administrations to

young and adult offenders were reported in 2010 (Wormith, 2011). The revised adult

version (LSI-R, Andrews & Bonta, 2000) is used in over 900 correctional agencies in

North America (Smith, Cullen & Latessa, 2009). The youth version and related

adaptations are used in 24 jurisdictions in Canada and the Unites States, and in 10

international locations (C. Tobias, personal communication, 16 December, 2014).

Substantial applied research over the past 30 years has found the Level of

Service (LS) inventories to have robust validity for predicting criminal offending

(Andrews & Bonta, 2010; Hoge & Andrews, 2011). A substantial number of primary

studies of the relationship between LS inventory scores and criminal recidivism have

been aggregated via meta-analyses for juvenile (e.g. Olver, Stockdale, & Wormith,

2009; Schwalbe, 2007) and adult offenders (e.g. Gendreau, Little, & Goggin, 1996).

Mean predictive validity effect sizes are typically in the range of $r = .25 - .40$

(Gendreau et al., 1996; Holtfreter & Cupp, 2007; Schwalbe, 2007). Upon this

foundation of sound validity and widespread utility, recent interest has focused on

improved understanding of sources of variability that may arise when the inventory is

used with different subgroups (e.g., by gender, race/ethnicity), with different offender

types (e.g., violent offenders, sex offenders), and in varied jurisdictions with different

quality control processes. This trend promises improvement by refining what exists

rather than proliferating new inventories (Campbell, French, & Gendreau, 2009;

Scurich & John, 2012). It is consistent with the advances in risk assessment and risk

management that are essential for integrated fourth generation risk assessment

(Andrews et al., 2006). The current research contributes to these developments. We

first investigated the predictive validity of the Australian Adaptation of the Youth

Level of Service/Case Management Inventory (YLS/CMI-AA; Hoge & Andrews,

1995) by gender and by race/ethnicity in a large sample of juvenile offenders. From

that sample, we then identified extreme cases of prediction error. We undertook a file

review of those cases for insights into the systems and processes that may have

undermined the validity and practical utility of the inventory in our sample.

Subgroup Differences in Validity of LS Inventories

Apart from consolidating predictive validity coefficients across primary

studies, meta-analysis can reveal sources of variation in effect size. For example,

aggregated effect sizes from primary studies with females may be compared with

meta-analytic results from studies that have primarily sampled males (e.g. Smith,

Cullen, & Latessa, 2009). Alternatively, within a given meta-analysis, potential

moderator variables can be coded to examine their association with validity indices

(Olver, Stockdale, & Wormith, 2014). In a meta-review of 40 systematic reviews and

meta-analyses incorporating 2,232 primary studies on forensic risk assessment from

1995 to 2009, the predictive validity of LS inventories is well represented (Singh and

Fazel, 2010). Six broad topics were identified that differentiate validity of risk assessment schemes by demographic characteristics (e.g., gender, ethnicity, age), and potential moderators (e.g., the definition of recidivism, type of offense, length of follow-up, and country in which a study was conducted).

One implication of Singh and Fazel's descriptive meta-review is that risk assessment validity is moderated by a variety of salient variables that deserve ongoing investigation. For example, the authors conclude that there is mixed and inconsistent evidence regarding equivalence in predictive validity by gender and ethnicity. Other authors take a more definitive stance on variability in predictive validity. For example, for both juvenile offenders (Schwalbe, 2008, p. 1377), and for the LSI-R with adult males and females (Smith, Cullen & Latessa, 2009), sufficient research supports the use of non-gendered risk assessment schemes for males and females. However, Andrews et al. (2012) observed higher predictive validity coefficients for females compared to males in an analysis of five different samples utilising both the LS/CMI and YLS/CMI. By contrast, others have taken the view that gender differences in risk assessment validity and pathways to crime are compelling. For instance Holtfreter and Cupp (2007) reviewed LSI-R research with female offenders and concluded that the inventory is not gender neutral. In particular, they argue that the inventory is deficient for assessing recidivism risk when gender specific needs and circumstances influence the criminal pathways of women. Shepherd, Luebbers, and Dolan (2012) support this gendered view and extend the argument to the moderating role of ethnicity.

The empirical literature on subgroup differences with LS inventories extends beyond the focus on predictive validity. Numerous studies have examined gender and ethnic identity differences on item, subscale and inventory total scores. Some of this

literature for youth versions of the LS instrument we summarised previously

(Thompson & McGrath, 2012), and concluded "…subscale and total inventory scores

on YLS inventories do differ by subgroup. However, there is considerable variability

in the direction and extent of differences depending on subgroup characteristics,

sample size, setting and jurisdiction" (p. 346). Our own research findings (Thompson

& McGrath, 2012, 2013), based on 3,568 youth under community supervision in New

South Wales, Australia showed some gender and ethnic differences. For example,

females scored higher than males on the overall score and on four of the domain

scores. Indigenous youth scored higher than the non-Indigenous Australians and other

ethnic groups on the overall score and on five domain scores. There were also various

differences at the item level between gender and ethnic subgroups. The item, domain,

and total score differences that we found varied in degree from negligible to

moderate. Subgroup differences at these levels do not necessarily imply instrument

bias, nor undermine predictive validity. However, such analyses are important to fully

elucidate and inform optimal use of the inventory across gender and ethnicity

Research examining subgroups differences has to date focussed primarily on

gender and race/ethnicity, but other categorical attributes of offenders such as age

groups, gang affiliation and neighbourhood type have also been examined (Chu,

Daffern, Thomas, & Lim, 2010; Olver, Stockdale, & Wong, 2012; Onifade, Petersen,

Bynum, & Davidson, 2011). The predictive validity of the instrument in relation to

various offense types has also been examined, notably for general and violent

recidivism as well as for sexual and non-sexual offenders (Caldwell & Dickinson,

2009; Schmidt, Sinclair, & Thomasdóttir, 2015; Viljoen, Elkovitch, Scalora, &

Ullman, 2009; Welsh, Schmidt, McKinnon, Chattha, & Meyers, 2008). Recent meta-

analytic findings have shown the instrument to be a better predictor of general ($r =$

.29) compared to violent ($r = 23$) recidivism, and a moderate predictor of sexual ($r =$

.11) recidivism (Olver et al., 2014).

Findings such as those summarized above have led to a broad, but by no

means unanimous, acceptance that the LS Inventories have similar, and practically

useful, validity with various subgroups across a range of different criminal justice

jurisdictions. Nevertheless, sufficient evidence exists to warrant ongoing research

attention to sources of variation. This is the prudent course of action for the applied

science of risk-need assessment and will benefit both the nomothetic and idiographic

pillars of offender assessment and treatment.

Other Source of Variability in Predictive Validity of LS Inventories

Apart from subgroup differences that may impact the validity of the LS

inventories, threats to the validity of risk/need assessment extend beyond such

categorical attributes of offenders. Predictive validity coefficients are estimates of the

theoretically true value of the relationship between a predictor and an outcome

variable. In practice, this relationship can be compromised by error in the

measurement of both. Consequently, precision in measurement is fundamental to

sound assessment practice, especially where there are important decisions that result

from that assessment (American Educational Research Association, 2014). Andrews

et al. (2011) re-iterate these points in relation to LS risk–need inventories. In addition,

they provide elaboration and analyses of sources of variability that may undermine the

integrity of predictor and criterion measure in LS validity studies. The first source of

variability identified was the practice of administering risk-need inventories. The

quality of these assessments may be influenced by factors such as the training, and

supervision of assessors, and the breadth and depth of information that informs the

assessment. In addition, related assessment protocols, policies and the agency work

culture may augment or dilute the accuracy of the inventory scores that are produced.

Another key source of variability concerns the integrity of the outcome measure and

that is typically a bifurcation of any future offending that is based on self-report or

official records. According to Andrews et al., additional moderators of predictive

validity across research studies arise from the impact of methodological procedures

such as the size and heterogeneity of the study sample and the length of the recidivism

follow-up. Andrews et al. subject their views about sources of variability to analyses

using several substantive data sets of LS validity studies. They conclude that length of

follow-up is well supported as a moderator. In addition, they found that researcher

affiliation and the region in which a validity study was conducted (Canadian versus

non-Canadian) were statistically associated with validity coefficients. They contend

that these variables were operating as proxies for procedural quality control over the

sources of variation and error that they believe operate across LS validity studies.

The present study

Drawing on the findings reviewed above, the present research investigated

sources of variation in the predictive validity of an established version of the Youth

Level of Service Case Management Inventory. This study both replicated and

extended the examination of subgroup differences reported (Thompson & McGrath,

2012; 2013) for the Australian Adaptation of the Youth Level of Service/Case

Management Inventory (YLS/CMI-AA; Hoge & Andrews, 1995). The present study

used a more recent and larger sample than the previous investigation. We extended

the previous research in two ways. First, we investigated predictive validity by

examining non-differentiated reoffending and also violent reoffending. In addition to

gender and ethnic subgroup comparisons, we broadly differentiated our sample

according to whether there was only one versus more than one YLS assessment on

record. Second, we built on the quantitative findings by undertaking a qualitative in-depth file review of selected cases within our sample for which the dichotomous reoffending outcome (recidivate versus not recidivate) was counter to the reasonable expectation based on the YLS/CMI-AA total score. Operationally, we referred to these cases as errors of prediction and aimed to understand the discrepancy between risk score and offending outcome. We expected this process would reveal factors associated with risk assessment and recidivism (persistence/desistance) that are informative about the predictive validity and usefulness of the inventory.

Method

Part 1: Subgroup differences in YLS/CMI-AA scores and predictive validity

YLS/CMI-AA.

The Australian adaptation of this youth risk-need inventory and its use by the NSW Department of Attorney General & Justice (Juvenile Justice) has been described by Thompson and Pope (2005) and Thompson & McGrath (2012, 2013). In brief, like the parent instrument, the adaptation includes risk/need items over eight domains. Possible scores for the domains vary from 3 to 9 and the cumulative total score possible is from 0-48

Data Set and Sample. In our 2011 study, we accessed YLS/CMI-AA inventory scores for 3568 juvenile offenders assessed in the years 2003 - 2005 inclusive. The data for the current study were accessed with appropriate institutional ethics approval from the same government departments and comprised all YLS/CMI-AA inventories for juvenile offenders completed in the 24-month period 1 July 2008 - 30 June 2010 inclusive. We obtained additional information such as juvenile custodial admissions, administrations of the YLS prior to the study time frame, and youth

involved in a trial multi-systemic regional program. This allowed us to refine the

sample and follow-up procedure as described next.

The data were restructured to link repeated assessments for the same

individuals yielding 4887 juvenile offenders. For the vast majority (91.6%), between

1 and 3 inventories were completed during the two-year target period. Our analysis

focused on the earliest chronological assessment in that period (subsequently referred

to as the index YLS). The data were screened to remove cases with invalid data ($n =$

4), cases in custody at the time of the assessment ($n = 312$), and cases of young

persons involved in the multi-systemic trial intervention ($n = 170$). The final sample

comprised YLS/CMI-AA inventory results for 4401 community-based juvenile

offenders. For 2890 (65.7%) of those youth, this was the first recorded administration

of the instrument. For the remainder of cases, ($n = 1511$; 34.3%), there was at least

one previous YLS assessment. The most recent previous assessment ranged from 19

to 2107 days (mean 328.38 days) prior to the index assessment. In short, for two

thirds of our sample, the index YLS was the first YLS and for one third of the sample

there was an historical YLS.

Age at assessment was distributed as 13.9% 14 years or under, 42.6% 15 or 16

years, and 43.6% 17 years or over. Mean age at assessment was 16.56 years ($SD =$

1.48). Males ($n = 3681$) made up 83.6% of the sample and females ($n = 720$) 16.4%.

Males ($M = 16.65$ years, $SD = 1.47$) were marginally older than the females ($M =$

16.34 years, $SD = 1.33$; $t_{4399} = -5.35$, $p < .001$, $r = .08$). Ethnic identity was recorded

by Juvenile Justice staff based on self-report by the young person or a family member.

Australian Indigenous status included youth identified as either Aboriginal or Torres

Strait Islanders. These individuals ($n = 1432$) comprised approximately one third

(34.3%) of the sample. Youth who were neither Indigenous nor from another ethnic

background were classified as Non-Indigenous Australian (n = 1916, 46%). Youth

having a non-Australian cultural background were classified as Australian Ethnic (n =

821, 19.7%). There were 73 different cultural backgrounds represented. The most

commonly reported were Lebanese, Maori (Indigenous New Zealanders), and

Samoan, with each making up approximately 10% of the Ethnic subgroup. For 5.3%

of the sample, ethnic/cultural status was unknown. There were small but significant

differences ($F_{2, 4166}$ = 55.22, p < .001, η^2 = .03) in age by ethnicity (Australian

Indigenous M = 16.29 years, SD = 1.61), Non-Indigenous (M = 16.70 years, SD =

1.35), Australian Ethnic (M = 16.89 years, SD = 1.32). We also found a small but

significant difference in mean age (16.4 years) for youth with the index YLS as their

first YLS and those youth (17.0 years) having had a previous YLS (t_{4399} = 14.71, p <

.001, r = .22).

  Recidivism. General recidivism was defined as any reoffense resulting in a

court conviction that occurred within one year of administration of the YLS/CMI-AA,

taking into account time spent in custody. Specifically, time spent in custody during

the year after administration of the YLS was added to the one-year follow-up period

to ensure equivalent time at risk across the sample. Time to reoffense was based on

the date of the earliest offense occurring after administration of the YLS rather than

the conviction date associated with that offense to eliminate variability resulting from

unequal legal processing times. The definition of violent recidivism was based on the

Australian Standard Offense Classification (Australian Bureau of Statistics, 2008),

which classifies offenses into 16 divisions that range from homicide to miscellaneous

offenses. Divisions 1 – 6 involve the infliction of some form of harm on another

person and offenses in these categories, specifically, homicide, assault, sexual assault,

abduction and robbery, formed the basis of our definition of a violent offense.

Reoffending data were obtained from the NSW Bureau of Crime Statistics and

Research, which maintains a database of all convictions (both adult and juvenile)

recorded in NSW since 1995. The database has been shown to be a reliable measure

of offending in NSW (Hua & Fitzgerald, 2006) but unfortunately does not extend

across Australia.

Part 2. File review of false negative and false positive YLS/CMI-AA predictions

In our sample of YLS/CMI-AA data, we identified 36 very low-risk cases

having a reoffense conviction within one year of the date of the index inventory. With

a YLS Total score between 1 and 4 inclusive, these cases were the lowest 2.2 % of

scores among all reoffenders in our overall sample. They represented 2.1% of all

cases (reoffenders and non-reoffenders) in that very low score range and the majority

(88%) did not reoffend. We also identified 29 very high-risk cases that did not have a

reoffense conviction within one year of the index inventory. In the overall sample of

non-reoffending youth, these cases (YLS Total score = 36-45 inclusive) were the

highest 1.1% of scores. Overall there were 70 youth in this very high-risk range, and

58.5% reoffended. Senior staff from Juvenile Justice (New South Wales) Department

of Attorney General tracked the storage location of hard copy paper files for 26 of the

36 low-risk cases and for 25 of the 29 high-risk cases. The files were relocated to the

department's head office and these fifty-one case files became the primary data that

we analyzed in situ.

Our overall strategy was to examine information in the files to acquaint

ourselves with each young person and the assessment process behind the YLS/CMI-

AA total score. In addition, we wanted to interrogate the discrepancy between the

YLS/CMI-AA score and the 1-year reoffending outcome. Logistically, we shared file

reviewing in a way that brought efficiency, consistency and methodological integrity to the process. This process evolved and the key elements were as follows.

1. All three investigators participated together in reviewing the first ten and last four files, but then worked in different pairings to review the other 37 files in batches of 14, 8 and 15 files respectively.

2. Most of the files contained standard documentation to which we directed our attention. This included police and court documents related to offending, YLS/CMI-AA inventory scores, and a detailed review (Background Report) of the young person's past and present life situation, character, habits, and offending behavior. Prepared by Juvenile Justice staff, the Background Report is based on an interview with the young person and relevant information from collateral sources. It informs sentencing decisions and is normally the precursor to completing the risk-need inventory and developing a case management plan. These plans, progress updates, and supervision notes were also part of the file documentation.

3. To process, distill and organize the varied and sometimes extensive file information, we focused on building a chronology of key events from the time of the index YLS/CMI-AA assessment and forward for at least one year. We paid particular attention to the Background Report preceding the index YLS and looked selectively at file material prior to that date in order to judge whether the risk-need assessment was consistent with available information. We examined file documentation following the date of the index inventory for insights into the 1 year offending or non-offending status of the case that we had previously determined through our predictive validity analyses. For a small number of cases, when information we considered important was not in the hard copy files, we were able to direct specific requests to senior Juvenile Justice staff with access to the Department's electronic file systems.

4. The time to work through each case varied but was on average approximately 45 minutes. In that time, we produced a one-page written summary of basic factual information and our formative impressions and interpretations.

5. Analytically, we worked to identify patterns in the data that were pertinent to the discrepancy between the index risk score and offending outcome. Our approach was broadly consistent with qualitative thematic analysis as described by Braun and Clarke (2006) in that it involved a systematic search for manifest consistencies and latent meaning across cases. This was both a distilling and interpretive process. Transparency and procedural detail are central to the trustworthiness of such qualitative methods (Braun & Clarke, 2006; Morrow, 2005). Serving this interest, but with regard to space limitations, we elaborate further on our analytic process.

6. We began to discuss insights and tentative data patterns in a piecemeal fashion as we worked through the file data and recorded our 1-page summaries. Our first attempt at enunciating these across the set of cases occurred when the three of us together completed the final four case reviews. These understandings were developed further over time. An oral presentation of our preliminary data patterns including case exemplars to senior Juvenile Justice staff engaged us further in our data and yielded some informed feedback. It was most important to ensure that our interpretive themes were grounded and supported by specific case file information and consistent across cases that we thought exemplified the patterns. Consequently, key information and initial interpretive comments from our one page case summaries were transferred into a tabular case by information matrix for the low-risk cases and for the high-risk cases. The conciseness of this data framework permitted a more thorough grasp of the most essential data. Specifically, this involved refining and checking the validity of commonalities across cases, and sorting of cases by common theme. The detailed

tabular analytic work was undertaken by one of the authors (AT) and then reviewed at

a general level for integrity and accuracy by the other authors.

Results

Part 1: Subgroup differences in YLS/CMI-AA scores and predictive validity

      Descriptive statistics. The mean and standard deviation for the YLS domain

and total scores are shown in Table 1 with breakdowns by gender and ethnicity. The

mean total YLS/CMI-AA score for the full sample was 15.11 ($SD = 8.30$). Females

scored higher on the instrument than the males ($t_{987} = 4.56$, $p < .001$, $r = .14$), and also

scored higher on the domain scores for Family and Living Circumstances, Substance

Abuse, Leisure and Recreation, and Personality/Behavior. The largest difference was

for the family and living circumstances scale, where the females scored just over half

a point higher than the males, which approached a moderate effect size ($t_{4167} = 8.50$, p

$< .001$, $r = .26$). In regard to ethnic/cultural identity, the Indigenous Australians had

the highest total score, followed by the Non-Indigenous Australians, with the

Australian Ethnic group scoring lowest (Welch's $F_{2, 2163} = 34.04$, $p < .001$, $r = .13$).

Post hoc ANOVA comparisons using the Games-Howell procedure were used to test

ethnic differences on domain scores. Significant domain differences were observed

for Prior and Current Offenses, Family and Living Circumstances, Peer Relations,

Substance Abuse, and Leisure and Recreation, with the Indigenous youth scoring

higher than the Ethnic youth on all five domains, and higher than the Non-Indigenous

youth on four but not Substance Abuse. Effect sizes ranged from .06 (Leisure and

Recreation) to .24 (Prior and Current Offenses). Comparing the subsets of our sample

having the index YLS as the first YLS with those having an historical YLS, there was

a difference in the YLS total score of 14.69 ($SD = 7.71$) versus 15.92 ($SD = 9.28$)

respectively. This difference is significant, although the effect size is small ($t_{4399} =$

4.68, p < .001, *r* = .07). For these two subsamples, we also note that there were five

significant domain scores differences. The largest was on Prior and Current Offenses

(t_{4399} = 26.28, p < .001, r = .37) with the First YLS subsample having a mean score 1.4

points lower.

Predictive validity. To begin with, we consider general recidivism, by which

we mean any reoffense, undifferentiated by the nature of the offense. Recidivism data

showed 19.3% (*n* = 848) of the sample reoffended within 3 months, 27.8% (*n* = 1224)

within 6 months and 37.4% (*n* = 1647) within a year. Mean time to the first reoffense

was 117.29 days (*SD* = 97.75). By subgroup, the one year general recidivism rates

were: Females 26.9% (*n* = 194), Males 39.5% (*n* = 1453), Indigenous 52.0% (*n* =

745), Non-Indigenous 32.0% (*n* = 613), and Ethnic 32.9% (*n* = 270). For the

subsample having their first YLS 35.5% (*n* = 1025) reoffended compared with 41.2%,

(*n* = 622) of those having an historical YLS prior to the index YLS.

The relationship between YLS/CMI-AA scores and general recidivism at one

year is shown in Table 2 for the full sample, and by gender and ethnic subgroups.

Point biserial correlations and receiver operating characteristic (ROC) curve analysis

are both provided as the latter are less susceptible to base rate variation than the

former (Babchishin & Helmus, 2016).

The point biserial correlation coefficient observed for the full sample was .31

and that represents a medium to large effect size for a base rate of 37.5% (Rice &

Harris, 2005). A coefficient of .30 was observed for females compared with .32 for

males and this gender difference was not statistically different. By ethnicity, point

biserial correlations ranged from .24 for the Indigenous group to .35 for the Ethnic

Australian group. The only significant difference was between the Indigenous and

Ethnic subgroups (*p* < .01). Apart from the predictive validity results for general

recidivism provided in Table 2, we note additionally that for general recidivism there

was a significant difference between the point biserial correlation for the First YLS

subsample (.28) versus the Historical YLS subsample ($r = .35$, $p = .014$).

AUC statistics in Table 2 show a significant value of .688 for the total sample.

The AUC differences by gender and ethnicity, as indicated by non-overlapping CIs,

mirrored the pattern of predictive validity correlation coefficients.

In addition to predictive validity for general recidivism, we examined the

relationship between YLS/CMI-AA scores and violent recidivism at one year. Overall

4.4% ($n = 194$) of the sample violently reoffended within 3 months, 6.6% ($n = 292$)

within 6 months and 9.3% ($n = 410$) within a year. The one-year violent reoffending

rates were 9.4% ($n = 68$) for females and 9.3% ($n = 342$) for males. By ethnicity

violent reoffending was 11.5% for Indigenous, 8% for Non-Indigenous, and 10.1%

for Ethnic youth. In order to examine predictive validity for violent offending, an

outcome variable was coded dichotomously where $0 = $ no offending in the follow-up

period ($n = 2754$) and $1 = $ violent offending ($n = 410$). This approach eliminated 1237

youth who reoffended non-violently from our analysis. In this reduced sample of

3164, the overall violent reoffending rate was 12.9%, and by subgroups, base rates

ranged from 10.6% for the Non-Indigenous subsample to 19.3% for the Indigenous

subsample. Predictive validity coefficients for violent recidivism by gender and

ethnicity are shown in the lower half of Table 2. The overall point biserial was $r = $

.20, and the overall AUC = .672. None of the subgroup correlation coefficients

differed statistically, and all of the associated AUC confidence intervals overlapped.

Apart from the predictive validity results for violent recidivism presented in Table 2,

we note parenthetically that the 1 year violent recidivism base rate for the subsample

having had an historical YLS was 41.2% versus 35.5% for the subsample with the

index YLS as the first YLS. The related violence predictive coefficients for these two subsamples were $r = .24$ and $r = .17$ respectively and these were not significantly different.

Part 2: File review of false negative and false positive YLS/CMI-AA predictions

Through the sequential qualitative analysis of file reviews described, we identified several explanatory patterns concerning the very low-risk reoffenders and a similar number of thematic consistencies for the very high-risk youth not reoffending. Some of these proto-themes were related at a broader interpretive level. The overarching and subordinate patterns are summarized in Table 3. Below, we describe these further with some brief exemplars to enhance the explanation and provide insight into the data underpinnings.

Reoffending Low risk youth. We reviewed 26 cases where youth with very low (1-4) risk/need scores reoffended within one year. All but one file related to a male offender. Seven were Indigenous, 10 Non-Indigenous, and eight Ethnic, with one case of unknown ethnic background. For the 26 cases reviewed, the most prominent meta-pattern was that the index YLS assessment underestimated to a degree the true extent or gravity of the young person's risks and needs. It was not possible to be definitive about the degree of suspected underestimate, but we had reason to believe it ranged from minimal to substantial. This was evident for 18 (69%) of the cases. Four scenarios supported our contention that the risk-need assessment was, to a degree, an underestimate. Three of these involved probable limitations to the collection, weighing, and synthesis of information about the young person being assessed. Specifically, for six cases we could not find on file the Background Report that should have been the foundation for completion of the inventory. The report may not have been written for various reasons, but the index YLS was nevertheless on file.

In one instance, the YLS total score was 4 but seven months later a new YLS was higher by 4 points and a related note stated that the initial YLS was "not true assessment". In a second case, the young person was difficult to contact and not attending interview appointments for a Background Report to be prepared. The index YLS seemed incomplete with only one item endorsed for a total score of 1. A third case without a documented predating Background Report received a YLS total score of 2, but four months later a new YLS inventory had a score of 14. For various reasons, three other YLS inventories without a predating Background Report raised the suspicion that risks and needs were underestimated. For one of these, the YLS total score was 4 but within the next year a subsequent YLS inventory score of 12 and a documented Background Report revealed multiple risks and needs of some duration.

The second scenario of underestimating risks and needs due to shortcomings in the supporting information involved five cases for which we were able to review the relevant predating Background Report, but it seemed to lack breadth and/or depth of information. Across the five cases, dynamic risk needs, such as those associated with family background, education, peers, alcohol and other drug use (AOD), and antisocial attitudes, were not as well represented or elaborated in the Background Report (BR) and YLS inventory as existing documentation or subsequent information revealed. For example, one youth and two co-accused were involved in a string of offenses but most of one youth's friends were described as "positive young people". In two cases, both with YLS scores of 3, subsequent information showed that the true extent of longstanding and heavy AOD use had not been disclosed and, in one case, unknown to family. For one of these cases, subsequent YLS assessments 7 and 8 months later had a total score of 28 and 26 respectively

The third underestimate scenario was evident in seven cases where the item endorsements for the index YLS were not congruent with pre-existing documentation on file. In these cases, a detailed and timely Background Report provided the informational underpinning for completion of the YLS but some of the static and dynamic risks reported or implied were not evident in item scores. In one case with a YLS total score of 3, we conservatively estimated domain scores based on the pre-dating Background Report and arrived at a total score of 10. This was in keeping with a total score of 11 from a subsequent YLS on file 14 months later. As another example, a documented trail of case notes and a Background Report told a vivid story of escalating crises and acting out at home, at school, and in the community by a young teen with a chronic childhood history of adversity and maladjustment. The index YLS total score of 4 was not consistent with this information and a subsequent YLS risk assessment 10 months later was ten points higher. Our judgment of underestimate was confirmed in another case with a YLS score total of 1. A subsequent file audit note stated that the YLS was incomplete and 4.5 months later a new assessment resulted in a total score of 23.

As noted, just over two thirds of the low risk files that we examined were accounted for by the three above scenarios. Of the remaining 8 cases, the low YLS total score seemed reasonably consistent with pertinent file documentation in five instances. For another three cases, we noticed a distinguishing feature that may have influenced the accuracy of the YLS. In these cases, the YLS total scores of 2, 3 and 4 had dropped 5, 10, and 16 points respectively from assessments 4 to 10 months earlier. In addition, the lower scores were temporally associated with an administrative transition point that involved a loosening of official community supervision of the young person. It seemed possible that a degree of optimism

associated with this process may have influenced the evaluation of risks and needs

beyond what we could see evidence for in the documentation.

High-risk youth not reoffending within one year. We reviewed 25 cases in

this category. By contrast with the low risk group, there were proportionately more

females ($n = 9$, 36%). In terms of ethnicity, the majority were Indigenous ($n = 13$),

followed by Non-Indigenous ($n = 8$) and then other ethnic backgrounds ($n = 4$). For a

large majority of the files reviewed, available information was consistent with a YLS

score in the high-risk category. For several cases, insufficient file documentation

made it difficult to confirm the high score, and for some cases we could not identify

evidence to support the extremely high YLS score. Sequential qualitative analysis of

the discrepancy between the high-risk scores and 1-year, non-offending status resulted

in several explanatory themes.

First, for nine cases, the non-offending status seemed artefactual due to

analytic and database blind spots. For example, three cases were spurious non-

offenders simply because we filtered our recidivism analysis based on a 1 year time

frame, but their reoffending occurred respectively 399, 444, and 547 days after the

index YLS. For the other 6 cases, there was credible file documentation of

reoffending within one year of the index YLS that was not lodged in the jurisdictional

database we accessed for follow-up. It was not entirely clear why this occurred but

information suggested the possibility that charges were not converted into

convictions, or there were delays in both conviction and entry into the database or,

quite clearly in one case, because the offense and conviction occurred in an adjoining

state jurisdiction.

A second pattern we identified also related to insights about the non-offending

outcome that were not evident at the quantitative, categorical level. We refer to the

nine cases evidencing this theme as pseudo non-offenders because various

circumstances seem to have prevented further offending in the 1-year follow-up

period or subsequently prevented it from being detected. For example, health or

mental health issues were likely mitigating or incapacitating factors in two cases. The

clearest example was an 18 year-old with a history of suicide ideation and suicide

attempts. Two months after his index YLS, he was in a mental health facility but

subsequently absconded. In spite of a warrant for his arrest, repeated file notes over

the next 3-4 years reported him as not contactable and whereabouts unknown. For

four of the high-risk cases, further offending was probably curtailed by varying

periods of detention that were not detectable in our data base screen. These cases

highlighted data base limitations, but also revealed the complexity of tracking

different types of detention such as remand, custody, or even periods of residential

treatment, especially for youth who move from the juvenile to adult justice system. In

addition, three other cases illuminated the itinerant lifestyle of some youth that can

obscure criminal activity. One youth moved inter-state to live with his girlfriend. The

family of another youth with an ethnic background may, according to a case note,

have sent their son for a "long holiday overseas". Another instance involved an 18

year-old female with many longstanding life difficulties and reportedly no safe and

secure accommodation several months after the index YLS. To complicate systemic

tracking of this young person, two different dates of birth were noted for her on file.

The two above patterns accounted for eighteen of the high-risk non-offending

youth, with nine cases representing each of the artefactual and pseudo non-offending

themes. Of the other seven case files that we reviewed, one showed some features of

both patterns. The remaining six cases were high-risk youth who based on available

file information appeared to have desisted from further offending during the follow-

up period. It was noteworthy that four of these were young females (aged 15 to 18 years at the time of the index YLS). It seemed likely that they had avoided further offending due to situational changes in their life circumstances and/or interventions relevant to the risks and needs typical of many females involved with the juvenile justice system. For example, risks and conflict in the family home had been precipitating factors for one young female's offending and these circumstances may have improved or changed as she approached 18 years of age several months after the index YLS. Other developmental changes that could have pre-empted further offending were exemplified in the case of a 15 year-old who became pregnant within months of the index YLS and, soon after, she was referred to health support services due to concerns about her unborn child. Two other high-risk females may have benefited from timely interventions relevant to their risks/needs. One was referred to an intensive therapeutic program unit with the juvenile justice system within a month of the index YLS. For the other female, it was noted 3 months after the index YLS that she was involved with community-based youth and family services. Finally, there were two indigenous high-risk males who also appeared to have avoided offending in the follow-up period, but we discerned no compelling insights into this outcome via the file review.

Discussion

This research using both quantitative and qualitative data pertinent to the predictive validity of the YLS/CMI-AA builds on our previous investigation into the predictive validity of the same inventory in the same juvenile justice jurisdiction (Thompson & McGrath, 2012, 2013). The sample for the current study was approximately 20% larger and the inventory data from a more recent time frame.

The quantitative analyses and results focussed on comparing YLS/CMI-AA

total and domain scores, as well as predictive validity, by gender and ethnic

subgroups in our sample. In broad terms, there were both similarities and modest

differences in mean scores across subgroups, but predictive validity indices remained

robust. Considering gender subgroups, females scored higher than males on the YLS

total score and on four of the eight domain scores (Family and Living, Substance

Abuse, Leisure and Recreation, Personality and Behaviour). The pattern, direction and

magnitude of these differences are consistent with the gender comparisons in the

earlier study (Thompson & McGrath, 2012, 2013). For example, the total score

difference was approximately 1.5 points higher for females in that sample and in this

one. The same four domains differed significantly, with the largest domain difference

being just over half a point higher for females on Family and Living Circumstances

previously, and in the current study. Such domain differences are in keeping with the

risk factors and pathways to crime that are thought to underlie female juvenile

offending (Dixon, Howie, & Starling, 2005; Reisig, Holtfreter, & Morash, 2006). In

this respect, the gender differences on YLS/CMI-AA domain and total scores may be

viewed as a re-assurance that the item content is reflective of what is known about

gender differences in juvenile offending. In our previous study, we compared item

endorsement proportions by subgroups to buttress this perspective, but did not drill

down to that level in the current research. Regardless, in the matter of potential

instrument bias for various subgroups, the flow on effect of mean score differences

must be weighed against predictive validity results which we summarize and discuss

below.

Comparisons between ethnic subgroups in our sample also showed various

differences on the YLS scores. The most prominent pattern was for Indigenous youth

to have higher mean scores than the Non-Indigenous subgroup and the Ethnic

subgroup. This was the case for the total score and for four of the eight domain scores

(i.e., Prior and Current Offenses, Family and Living Circumstances, Peer Relations,

Leisure and Recreation). The Ethnic subgroup had the lowest YLS total score and the

lowest mean score on three domains (Prior and Current Offenses, Family and Living

Circumstances, Substance Abuse, Leisure and Recreation). Most of these Ethnic

subgroup differences were small, but they overlap largely with the same findings in

our previous study (Thompson & McGrath, 2012, 2013). In the Australian context,

the pattern for Indigenous youth is consistent with well-documented manifestations of

past and ongoing cultural disadvantage (Bradley, Draca, Green, & Leeves, 2007;

Cooke, Mitrou, Lawrence, Guimond, & Beavon, 2007). The indications of small, but

significantly lower, criminological risk for Ethnic youth as a group is noteworthy and

deserving of further, more granular investigation, especially given widespread

negative discourses surrounding immigrants and refugees (Schweitzer, Perkoulidis,

Krome, Ludlow, & Ryan, 2005).

       The pattern of predictive validity results for general recidivism in the current

study was similar to findings in our previous study, although the magnitude of the

validity indices improved to a degree. For instance, we previously observed an overall

point biserial correlation of .26 and an AUC of .652. In the current study, those

indices were significantly higher at .31 and .688 respectively. By gender and ethnic

subgroup, point biserial correlations in our earlier study ranged from .17 to .27 and

AUCs from .604 to .659. In the current study, those ranges were .24 to .35 and .648 to

.716 respectively. These increases in predictive validity indices may be due to

improvements in assessment practices as well as our research methodology over time.

For instance, departmental use of the YLS/CMI-AA has become better established

since the time-frame of the previous study and we were able to adjust for time spent in

custody during the follow up period but previously did not access that information.

Overall, the pattern of predictive validity results previously, and in the current study,

showed comparable and not significantly different indices by gender. The same was

largely the case for predictive validity by ethnic subgroups. Hence, these findings

support the use of YLS/CMI-AA across various subgroups but it is prudent and useful

to monitor YLS performance characteristics for the diversity of juvenile offenders in

any jurisdiction. For example, we have argued and demonstrated (Thompson &

McGrath, 2012) that small differences in predictive validity coefficients and base

rates can lead to subgroup difference in predictive accuracy when score bands rather

that YLS total scores are used to categorically assign levels of risk/need.

We found in the current study the YLS/CMI-AA was a significant predictor of

violent offending and there were no subgroup differences in the validity coefficients.

Our results are in line with those reported in the international literature. For instance,

Olver et al. (2014) reported a random effects size for violent recidivism of .25 for

females and .29 for males while we observed a point biserial correlation of .26 for

females and .19 for males.

However, we draw attention to criterion definition when comparing predictive

validity for violent recidivism across studies. The dichotomous criterion that we used

was "violent reoffense" versus "no reoffense". We are not aware of any consensus

about procedural and outcome measures in the violent recidivism literature, but

suspect that considerable variability exists. To emphasize this point, we note that the

logic of our operational procedure excluded from analysis youth who committed a

reoffense that was not violent. It could be argued that an alternative criterion of

violent reoffending versus other outcomes (i.e., desistance or non-violent recidivism)

is also conceptually and pragmatically beneficial if, for example, one were trying to find predictors that would help identify youth most in need of an intervention program to reduce violent recidivism. Base rates are also affected by these criterion considerations. Using the alternative operational definition (violent recidivism versus desistance/non-violent recidivism) with our data changed the base rate from 12.9% to 9.3% and the validity coefficients (point biserial and AUC) from .20 and .672 to .10 and .611 respectively). Thus, it is important to be clear about the conceptual, applied, and statistical implications of how violent recidivism is operationally defined.

Finally, a comment is on the differentiation of our data according to whether the index YLS was the first YLS as opposed to there being an antecedent YLS assessment. The descriptive and predictive analyses comparing these subsets of data revealed small but significant difference in YLS domain and total scores and in validity coefficients for general recidivism. As suggested by the related results that we presented, this may have been due to youth with an antecedent YLS being an older and historically higher risk cohort. These findings are intriguing because they show that an admittedly gross attempt at looking at YLS score relativity may have salutary benefits for improving predictive validity. As many of the youth in our overall sample had more than one YLS assessment in the two-year target time frame, it would have been illuminating to examine the predictive validity of change in contiguous or fixed interval risk scores. We did not pursue this but believe it is consistent with the expectation that, in best practice, there should be timely repeat assessments. Also, some research on this topic has been conducted (e.g., Labrecque, Smith, Lovins, and Latessa, 2014, and Vose, Lowenkamp, Smith, and Cullen, 2009 have examined changes in YLS scores as predictors of recidivism).

Thus far we have discussed the quantitative findings of our research but we also undertook a qualitative investigation of what we referred to as 'invalid' (i.e., intuitively incongruent instances) of YLS total score and one-year reoffending outcome. Specifically, we examined case file and YLS documentation for a sample of young persons who had extremely low YLS total scores but nevertheless reoffended, and a similar number of high scoring persons who did not reoffend. The collection and analysis of that data resulted in a number of broad and subordinate explanatory themes. We highlight and deliberate on those findings next.

Foremost, the essential thread of the qualitative results is that there are plausible and factually grounded explanations for a large majority of the cases that were associated with discrepant reoffending outcomes. For the low risk cases, the most compelling interpretation of the broad and subordinate themes was that procedural aspects of YLS assessment undermined validity. For approximately 70% of the cases, validity was likely compromised by limitations and shortcomings in the collection and integration of pertinent information or in converting information into high fidelity risk/need item endorsements. For such reasons, and based on available documentation, most of these cases seemed to underestimate actual risk/need. Although, we could not be definite about the degree of underestimate, the major outcome was that the reoffending outcome for our very low risk sample was not as glaring or unexpected as the surface level data implied. In this regard, these results are consistent with the two general principles of YLS administration guidelines (Hoge & Andrews, 2011, p. 5). Specifically, sound assessment rests on the pillars of the best available information and sound professional judgment.

As previously noted, deficiencies in these underpinnings is one of the key sources of variation in YLS validity across studies according to Andrews et al. (2011).

Those authors also point out that understanding such sources can inform improved procedural integrity in field settings. The qualitative procedure that we adopted in the current research is useful in achieving that aim because it provided context specific practices and vivid case examples that should be particularly useful for training and quality assurance as a penetrating supplement to administration guidelines and general exhortations about integrity of the predictor measure. Beyond this, the thematic analysis of the very low risk reoffending cases brought to light the possibility of judgment dynamics such as positivity or confirmation biases that may arise, for example, when YLS assessment is linked to administrative decisions points for loosened supervision. The limits of human information processing including heuristic and judgmental biases are well known (e.g. Kahneman & Tversky, 1996; West, Meserve, & Stanovich, 2012) and it may prove fruitful to relate and incorporate some of the key lessons of that literature to the cognitive demands of YLS assessment.

For high risk, non-offending cases, the crux of the qualitative results was that the high risk/need total score was justified, but for approximately three quarters of the cases, the follow-up, non-offending status was open to question. In essence, the major reasons for doubting that these youth desisted from offending was because of evidence to the contrary or due to factors that circumvented 1 year recidivism. Collectively, these explanatory themes are representative of most of the criterion and methodology issues that Andrews et al (2011) emphasize as sources of variability in LS validity estimates. Beyond this though, the current study revealed jurisdiction-specific blind spots in tracking recidivism and exemplars not mentioned by Andrews et al., such as such as various forms of incapacitation and lifestyle instability that may foil follow-up tracking.

In summary, for the sample of case files that we reviewed, inaccuracies in the predictor and outcome measures accounted, in large part, for the disjunction between YLS total score and one-year recidivism. In itself, this conclusion is not surprising, but the qualitative methodology and the detailed nature of the findings does constitute a novel and useful approach. In relation to this, there are additional points that deserve comment.

First, we draw attention to those cases in our qualitative analysis for which: (a) the YLS total score seemed justified based on available documentation, and (b) we had no reason to doubt the recidivism outcome. Specifically, there were five very low-risk youth who reoffended and six very high-risk youth who did not reoffend. These cases are the best examples of putative false negative and false positive YLS predictions, and they afforded an opportunity to explore and speculate on the dynamics. The clearest insight was the role of developmental events and timely interventions that may have reduced the criminological risk for four of the high-risk females (final subordinate theme in Table 3). We gained little by way of understanding why the legitimately low risk individuals reoffended, except to note that two were in fractious (high risk) family situations that potentially facilitated offending. A second point to address concerns the relationship between the qualitative and quantitative findings of the current research. The sample for qualitative analysis constituted just over 1% of the quantitative sample. Even if we could have rectified the shortcomings of the predictor and outcome variables in that small subsample, we think it would not likely have impacted significantly on the reported validity indices in Table 2. Unknown, however, is the extent to which the kind of limitations we identified may have infiltrated the very large sample of remaining cases. It is reasonable to conclude that the validity indices underestimate to some extent,

although not necessarily a significant degree, true validity in our sample. Furthermore,
the cases that we referred to as the best exemplars of false negatives and false
positives are valuable reminders for researchers and practitioners that prediction of
juvenile offending will always be qualified by the fluid and elusive nature of
interaction between individual and context, including circumstances and events that
may facilitate or prevent criminal behavior.

There are limitations to the research that we have reported here. In essence,
the qualitative analysis provided a detailed account of shortcomings in our
quantitative investigation into YLS/CMI-AA validity. As noted previously, such
limitations are familiar. However, readers may be less familiar with indicators for
assessing the reliability and validity of the qualitative component of our research. A
full account of standards for assessing rigor in qualitative research (see for example,
Morrow, 2005) is beyond the scope of this research report. However, the following
are important considerations. The detailed account that we provided of our method
and analysis is fundamental to judgments about the sufficiency and thoroughness of
our procedures. In addition, case data excerpts and links to the risk-need literature
have been provided to support both the internal integrity and external validity of the
themes that we derived. Researcher subjectivity and bias are endemic to all research
and the impact of such influences on our qualitative procedures is a moot point. Á
propos, while reviewing case documentation and in the subsequent analyses, we were
not explicitly seeking to exonerate the YLS/CMI-AA from prediction errors, nor did
we have prior interpretive expectations for sources of error variability based on
Andrews et al (2011) which we had not yet read. The crux of our procedure though
was to examine cases that represented errors of prediction and in subtle ways that
could bring with it a perspective different from examining, at a similar level of detail,

YLS assessments and being blind to the reoffending outcome. Although we three authors shared equivalently in the case review data collection, we do acknowledge different degrees of immersion in that data at the case coding and theme development stage. Hence, we could have engaged in a more equivalently detailed process of co-analysis to construct and reliably confirm meaning in the raw data. Qualitative methods often make an explicit effort to engage participants often marginalized or overlooked by other research methodologies. Considering this point, it is important to acknowledge that our method did not give voice to two sources that may contribute much to the kind of understandings we were seeking. Specifically, the Juvenile Justice Officers who completed the risk assessment inventories, and the young people who either reoffended or did not reoffend, must be recognized as highly valuable informants for an even more thorough and triangulated analysis than we have provided.

In conclusion, this research has provided predictive validity indices for the YLS/CMI-AA based on a large Australian sample and differentiated by some of the key variables recognized as important in the risk assessment literature on juvenile offending. The qualitative component of the research was novel and complementary. It yielded interesting findings that are valuable for various reasons. Most importantly, the qualitative results may be said to have consequential validity. In other words, those findings have transparent and concrete implications for actions that may be undertaken to optimize predictive validity by individuals or systems involved in risk-need assessment with the YLS/CMI-AA.

Author biographies

Andrew McGrath lectures in forensic and developmental psychology at Charles Sturt University in Bathurst, Australia. He received his PhD from the University of Sydney in 2007. His research interests are in juvenile justice, criminological theory, risk assessment and bullying.

Anthony Thompson completed his PhD in Psychology at the University of Western Ontario in 1975. Since that time he has worked as a registered psychologist and as an academic in Canada and Australia. He is now fully retired and devotes himself to honoring everyday life.

Trained in law and experimental psychology, Jane Goodman-Delahunty is a Research Professor at Charles Sturt University and a Member of the New South Wales Civil and Administrative Tribunal. An elected Fellow of the American Psychological Association, she is a former editor of Psychology, Public Policy & Law, and served for 10 years on the New South Wales Law Reform Commission.

References

American Educational Research Association, American Psychological Association,

National Council on Measurement in Education. (2014). *Standards for*

educational and psychological testing. Washington DC: American

Educational Research Association.

Andrews, D. A., & Bonta, J. (2000). *The Level of Service Inventory-Revised*. Toronto:

Multi-Health Systems.

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.).

New Providence: Anderson Publishers.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2004). *Level of Supervision/Case*

Management Inventory (LSI/CMI): An offender assessment system. User's

guide. Toronto: Multi-Health Systems.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of

risk and/or need assessment. *Crime & Delinquency, 52*(1), 7-27.

Andrews, D. A., Bonta, J., Wormith, J. S., Guzzo, L., Brews, A., Rettinger, J., &

Rowe, R. C. (2011). Sources of variability in estimates of predictive validity:

A specification with level of service general risk and need. *Criminal Justice*

and Behavior, 38, 413-432. doi:10.1177/0093854811401990

Andrews, D. A., Guzzo, L., Raynor, P., Rowe, R. C., Rettinger, L. J., Brews, A., &

Wormith, J. S. (2012). Are the major risk/need factors predictive of both

female and male reoffending?: A test with the eight domains of the Level of

Service/Case Management Inventory. *International Journal of Offender*

Therapy and Comparative Criminology, 56, 113-133.

doi:10.1177/0306624x10395716

Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P., & Cullen, F. T.

(1990). Does correctional treatment work? A clinically relevant and

psychologically informed meta-analysis. *Criminology, 28*, 369-404.

Australian Bureau of Statistics. (2008). Australian Standard Offense Classification

(ASOC). Canberra: Australian Bureau of Statistics.

Babchishin, K. M., & Helmus, L.-M. (2016). The influence of base rates on

correlations: An evaluation of proposed alternative effect sizes with real-world

data. *Behavior Research Methods, 48*, 1021-1031. doi:10.3758/s13428-015-

0627-7

Bradley, S., Draca, M., Green, C., & Leeves, G. (2007). The magnitude of educational

disadvantage of indigenous minority groups in Australia. *Journal of

Population Economics, 20*, 547-569. doi:10.1007/s00148-006-0076-9

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative

research in psychology, 3*, 77-101.

Caldwell, M. F., & Dickinson, C. (2009). Sex offender registration and recidivism

risk in juvenile sexual offenders. *Behavioral Sciences & the Law, 27*, 941-956.

doi:10.1002/bsl.907

Campbell, M. A., French, S., & Gendreau, P. (2009). The prediction of violence in

adult offenders. *Criminal Justice and Behavior, 36*, 567-590.

doi:10.1177/0093854809333610

Chu, C. M., Daffern, M., Thomas, S., & Lim, J. Y. (2010). Violence risk and gang

affiliation in youth offenders: a recidivism study. *Psychology, Crime & Law,

18*, 299-315. doi:10.1080/1068316x.2010.481626

Cooke, M., Mitrou, F., Lawrence, D., Guimond, E., & Beavon, D. (2007). Indigenous

well-being in four countries: An application of the UNDP's Human

Development Index to Indigenous Peoples in Australia, Canada, New Zealand,

and the United States. *BMC International Health and Human Rights, 7*, 9.

doi:10.1186/1472-698X-7-9

Dixon, A., Howie, P., & Starling, J. (2005). Trauma exposure, posttraumatic stress,

and psychiatric comorbidity in female juvenile offenders. *Journal of the

American Academy of Child & Adolescent Psychiatry, 44*, 798-806.

doi:10.1097/01.chi.0000164590.48318.9c

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of

adult offender recidivism: What works! *Criminology, 34*, 575-607.

Hoge, R. D., & Andrews, D. A. (1995). *Australian Adaptation of the Youth Level of

Service/Case Management Inventory*. North Tonawanda: NY: Multi-Health

Systems.

Hoge, R. D., & Andrews, D. A. (2011). *YLS/CMI 2.0. Youth Level of Service/Case

Management Inventory 2.0*. North Tonawanda, NY: Multi-Health Systems.

Holtfreter, K., & Cupp, R. (2007). Gender and risk assessment. *Journal of

Contemporary Criminal Justice, 23*, 363-382.

doi:10.1177/1043986207309436

Hua, J., & Fitzgerald, J. (2006). Matching court records to measure reoffending *Crime

and Justice Bulletin: Contemporary issues in crime and justice* (Vol. No. 95).

Sydney: NSW Bureau of Crime Statistics and Research.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions.

Psychological Review, 103, 582-591.

Labrecque, R. M., Smith, P., Lovins, B. K., & Latessa, E. J. (2014). The importance

of reassessment: How changes in the LSI-R risk score can improve the

prediction of recidivism. *Journal of Offender Rehabilitation, 53*, 116-128.

doi:10.1080/10509674.2013.868389

Morrow, S. L. (2005). Quality and trustworthiness in qualitative research in

counseling psychology. *Journal of Counseling Psychology Special Issue:
Knowledge in Context: Qualitative Methods in Counseling Psychology
Research, 52*, 250-260.

Olver, M. E., Stockdale, K. C., & Wong, S. C. (2012). Short and long-term prediction

of recidivism using the Youth Level of Service/Case Management Inventory

in a sample of serious young offenders. *Law & Human Behavior, 36*, 331-344.

doi:10.1037/h0093927

Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2009). Risk assessment with young

offenders: A meta-analysis of three assessment measures. *Criminal Justice
and Behavior, 36*, 329-353. doi:10.1177/0093854809331457

Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2014). Thirty years of research on

the Level of Service scales: a meta-analytic examination of predictive

accuracy and sources of variability. *Psychological Assessment, 26*, 156-176.

doi:10.1037/a0035080

Onifade, E., Petersen, J., Bynum, T. S., & Davidson, W. S. (2011). Multilevel

recidivism prediction: Incorporating neighborhood socioeconomic ecology in

Juvenile Justice risk assessment. *Criminal Justice and Behavior, 38*, 840-853.

doi:10.1177/0093854811407026

Reisig, M. D., Holtfreter, K., & Morash, M. (2006). Assessing recidivism risk across

female pathways to crime. *Justice Quarterly, 23*, 384-405.

doi:10.1080/07418820600869152

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies:

ROC Area, Cohen's d, and r. *Law and Human Behavior, 29*, 615-620.

doi:10.1007/s10979-005-6832-7

Schmidt, F., Sinclair, S. M., & Thomasdóttir, S. (2015). Predictive validity of the

Youth Level of Service/Case Management Inventory with youth who have

committed sexual and non-sexual offenses: The utility of professional

override. *Criminal Justice and Behavior, 43*, 413-430.

doi:10.1177/0093854815603389

Schwalbe, C. S. (2007). Risk assessment for juvenile justice: A meta-analysis. *Law &

Human Behavior, 31*, 449-462. doi:10.1007/s10979-006-9071-7

Schwalbe, C. S. (2008). A meta-analysis of juvenile justice risk assessment

instruments. *Criminal Justice and Behavior, 35*, 1367-1381.

doi:doi:10.1177/0093854808324377

Schweitzer, R., Perkoulidis, S., Krome, S., Ludlow, C., & Ryan, M. (2005). Attitudes

towards refugees: The dark side of prejudice in Australia. *Australian Journal

of Psychology, 57*, 170-179. doi:10.1080/00049530500125199

Scurich, N., & John, R. S. (2012). Prescriptive approaches to communicating the risk

of violence in actuarial risk assessment. *Psychology, Public Policy, & Law,

18*, 50-78. doi:10.1037/a0024592

Shepherd, S. M., Luebbers, S., & Dolan, M. (2012). Gender and ethnicity in juvenile

risk assessment. *Criminal Justice and Behavior*, *40*, 388-408.

doi:10.1177/0093854812456776

Singh, J. P., & Fazel, S. (2010). Forensic risk assessment. *Criminal Justice and

Behavior, 37*, 965-988. doi:10.1177/0093854810374274

Smith, P., Cullen, F. T., & Latessa, E. J. (2009). Can 14,737 women be wrong? A

meta-analysis of the LSI-R and recidivism for female offenders. *Criminology

& Public Policy, 8*, 183-208. doi:10.1111/j.1745-9133.2009.00551.x

Thompson, A. P., & McGrath, A. (2012). Subgroup differences and implications for

contemporary risk-need assessment with juvenile offenders. *Law and Human

Behavior, 36*, 345-355. doi:10.1037/h0093930

Thompson, A. P., & McGrath, A. (2013). Correction to Thompson and McGrath

(2012). *Law & Human Behavior, 37*, 106.

Thompson, A. P., & 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*, 207-214.

Viljoen, J. L., Elkovitch, N., Scalora, M. J., & Ullman, D. (2009). Assessment of

reoffense risk in adolescents who have committed sexual offenses: Predictive

validity of the ERASOR, PCL:YV, YLS/CMI, and Static-99. *Criminal Justice

and Behavior, 36*, 981-1000. doi:10.1177/0093854809340991

Vose, B., Lowenkamp, C. T., Smith, P., & Cullen, F. T. (2009). Gender and the

predictive validity of the LSI-R. *Journal of Contemporary Criminal Justice,

25*, 459-471. doi:10.1177/1043986209344797

Welsh, J. L., Schmidt, F., McKinnon, L., Chattha, H. K., & Meyers, J. R. (2008). A

comparative study of adolescent risk assessment instruments: Predictive and

incremental validity. *Assessment, 15*, 104-115.

doi:10.1177/1073191107307966

West, R. F., Meserve, R. J., & Stanovich, K. E. (2012). Cognitive sophistication does

not attenuate the bias blind spot. *Journal of personality and social psychology,*

103, 506-519. doi:0.1037/a0028857

Table 1

Mean (standard deviation) for the YLS/CMI-AA scores by gender and by ethnic subgroups

YLS/CMI-AA domain	Gender		Australian Ethnicity		
	Female	Male	Indigenous	Non-Indigenous	Ethnic
	($n = 720$)	($n = 3681$)	($n = 1432$)	($n = 1916$)	($n = 821$)
Prior and Current offenses	3.51 (1.62)	3.45 (1.80)	4.08_a (1.86)	3.27_b (1.68)	3.07_c (1.62)**
Family and living	2.41 (1.73)	1.81 (1.65)**	2.20_a (1.68)	1.87_b (1.68)	1.67_c (1.63)**
Education/employment	2.02 (1.85)	2.03 (1.89)	2.13 (1.80)	2.02 (1.92)	1.95 (1.98)
Peer relations	1.89 (1.32)	1.86 (1.27)	2.20_a (1.25)	1.67_b (1.26)	1.87_c (1.26)**
Substance abuse	2.18 (1.86)	1.98 (1.82)*	2.19_a (1.88)	2.10_a (1.84)	1.65_b (1.70)**
Leisure and Recreation	1.37 (1.11)	1.20 (1.10)**	1.33_a (1.10)	1.22_b (1.12)	1.13_b (1.07)**
Personality/behavior	2.17 (2.03)	1.73 (1.84)**	1.75 (1.85)	1.88 (1.91)	1.82 (1.88)
Attitudes/beliefs	0.91 (1.30)	0.80 (1.19)	0.84 (1.22)	0.82 (1.21)	0.85 (1.26)
Total score	16.45 (8.67)	14.85 (8.20)**	16.71_a (8.20)	14.86_b (8.34)	14.02_c (8.18)**

Note. For ethnic subgroups, means with different subscripts differ significantly. *$p < .01$. **$p < .001$.

Table 2: One Year Predictive Validity for YLS/CMI-AA Total Score by Subgroups

			General recidivism		
			ROC Analysis		
Group	*n*	*r*	*AUC*	*SE*	95% CI [LL, UL]
Female	720	.30**	.690**	.022	[.647, .733]
Male	3681	.32**	.694**	.009	[.677, .711]
Indigenous	1432	.24**	.648**	.014	[.620, .676]
Non-Indigenous	1916	.30**	.684**	.013	[.659, .708]
Ethnic	821	.35**	.716**	.019	[.680, .753]
Total	4401	.31**	.688**	.008	[.672, .703]
			Violent recidivism		
Female	594	.26**	.725**	.031	[.664, .786]
Male	2570	.19**	.667**	.015	[.637, .696]
Indigenous	851	.16**	.623**	.024	[.577, .670]
Non-Indigenous	1457	.18**	.668**	.022	[.626, .771]
Ethnic	634	.25**	.715**	.028	[.660, .771]
Total	3164	.20**	.672**	.014	[.646, .699]

Note: violent recidivism coded 0 = non reoffender; 1 = violent reoffender (non-violent reoffenders excluded)
** *p* < .001

Table 3: Summary of Themes from Qualitative Analysis of Case File Reviews (*N* = 51)

Very Low Risk Youth Reoffending (*n* = 26)

Broad Theme	Subordinate Theme
YLS Underestimates Risks/Needs (*n* = 18)	
	> Background Report Missing (*n* = 6)
	> Background Report lacks depth/breadth (*n* = 5)
	> Items on YLS inconsistent with file information (*n* = 7)
YLS consistent with file information (*n* = 5)	
YLS completed for administrative transition (*n* = 3)	

Very High Risk Youth not Reoffending (*n* = 25)

Broad Theme	Subordinate Theme
Non-offending status artefactual (*n* = 9)	
	> Reoffended after 1-year time frame (*n* = 3)
	> Credible file documentation of 1-year recidivism not showing in the follow-up database (*n* = 6)
Pseudo non-offending due to incapacitation or difficulty tracking (*n* = 9)	
	> Health or mental health issues (*n* = 2)
	> Periods of detention/custody not picked up in the database screen (*n* = 4)
	> Itinerant lifestyle (*n* = 3)
YLS consistent with file information (*n* = 6)	> Risk reduced for females by life changes and interventions (*n* = 4)

Note. One of the 25 high-risk cases showed features of both the artefactual and pseudo offending themes.