It is the paper published as:

Author: A. P. Thompson and A. McGrath
Title: Subgroup Differences and Implications for Contemporary Risk-Need Assessment with Juvenile Offenders
Journal: Law and Human Behavior **ISSN**: 0147-7307 1573-661X
Year: 2012 **Volume:** 36 **Issue:** 4 **Pages:** 345-355

Abstract: Risk-need assessment is widely accepted as best practice with juvenile offenders and is underpinned by a healthy research literature on risk assessment inventories. Previous studies have found both similarities and differences on risk measures when gender and racial/ethnic subgroups have been compared. Differential validity has been examined, but differential prediction has been overlooked. The current study undertook gender and ethnic comparisons for a large sample (N = 3568) of community based juvenile offenders who were evaluated using the Australian Adaptation of the Youth Level of Service/Case Management Inventory (YLS/CM-AA). Analyses showed various gender and ethnic differences at the item level, across domain scores and on the total inventory score, but not for validity indices. However, 1 year re-offending rates for youth in three classification categories (low, moderate, high) varied by gender and ethnicity. The findings were related to contemporary understandings of the risk factors for offending and the dynamics of crime for gender/ethnic subgroups. It is argued that in spite of these subgroup differences, a generic inventory such as the YLS/CM-AA can be used fairly with various subgroups. Recommendations for how this could be accomplished are provided.

Author Address: athompson@csu.edu.au
amcgrath@csu.edu.au

Subgroup Differences and Implications for Contemporary Risk-Need Assessment with Juvenile Offenders

Anthony P. Thompson

&

Andrew McGrath

Charles Sturt University

Author Note

Drs. Anthony Thompson and Andrew McGrath, Department of Psychology, New South Wales, Charles Sturt University, Australia.

The authors thank the NSW Government Department of Attorney General & Justice (Juvenile Justice) and the NSW Bureau of Crime Statistics and Research for assistance in undertaking the research. The opinions here do not necessarily reflect the views of these organizations, or any of their officers.

Some of the results were reported as a referred conference paper. In M. Katsikitis (Ed.), P*roceedings of the 2006 Joint Conference of the APS and NZPsS* (pp.445-449). Melbourne, Victoria, Australia: The Australian Psychological Society Ltd.

Correspondence concerning this article should be addressed to Dr. Anthony Thompson, School of Psychology, Charles Sturt University, Locked Bag 588, Wagga Wagga, NSW, 2678, Australia.

Key words: Juvenile Offending, Risk Assessment, Youth Level of Service Inventory.

Abstract

Risk-need assessment is widely accepted as best practice with juvenile offenders and is underpinned by a healthy research literature on risk assessment inventories. Previous studies have found both similarities and differences on risk measures when gender and racial/ethnic subgroups have been compared. Differential validity has been examined, but differential prediction has been overlooked. The current study undertook gender and ethnic comparisons for a large sample (N = 3568) of community based juvenile offenders who were evaluated using the Australian Adaptation of the Youth Level of Service/Case Management Inventory (YLS/CM-AA). Analyses showed various gender and ethnic differences at the item level, across domain scores and on the total inventory score, but not for validity indices. However, 1 year re-offending rates for youth in three classification categories (low, moderate, high) varied by gender and ethnicity. The findings were related to contemporary understandings of the risk factors for offending and the dynamics of crime for gender/ethnic subgroups. It is argued that in spite of these subgroup differences, a generic inventory such as the YLS/CM-AA can be used fairly with various subgroups. Recommendations for how this could be accomplished are provided.

Subgroup Differences and Implications for Contemporary Risk-Need Assessment with Juvenile Offenders

Third generation assessment of risks and needs is well established in both adult and juvenile corrections (Andrews, Bonta, & Wormith, 2006; Hoge, 2008; Schwalbe, 2008). For juvenile offenders, numerous risk-need inventories have been developed (Baker, Jones, Roberts, & Merrington, 2003; Howell, 1995; Thompson & Putniņš, 2003). The literature detailing the implementation and psychometric properties of these young offender inventories has grown substantially over the past 20 years. For example, sufficient empirical studies of structured inventories and their predictive validity exist for meta-analyses to be conducted (Olver, Stockdale, & Wormith, 2009; Schwalbe, 2007). One of the widely used risk-need inventories is the Youth Level of Service/Case Management Inventory (YLS/CMI; Hoge & Andrews, 2006) and its associated adaptations (Olver et al., 2009). A healthy research literature exists to support the YLS family of inventories. However, the psychometric properties for diverse subgroups of offenders is explored less often and is limited in depth of analysis. Olver et al.(2009) found that less than half of the YLS studies reported predictive validity indices for male versus female subsamples and only a small subset of Canadian studies compared Aboriginal with non-Aboriginal youth.

There are several reasons why risk-need assesment should be examined at the subgroup level. First, it is likely that risk factor profiles vary by subgroups. For example, comorbidty of problems, including exposure to trauma and victimisation, have been identified as more typical of female than male juvenile offenders (Gavazzi, Yarcheck, & Chesney-Lind, 2006; Mullins, Cornille, Mullins, & Huber, 2004; Odgers, Moretti, & Reppucci, 2005). Indigenous compared with non-Indigenous offenders are more likely to experience a confluence of psychological, social and environmental risks due to the evolving impact of colonisation and economic disadvantage (Austin, 2010; Cunneen & White, 2007;

Pedersen, Dudgeon, Watt, & Griffiths, 2006). Discrepancies in juvenile arrest rates for Black versus White American youth have been linked to increased levels of risks in early childhood such as conduct problems, low academic achievement, poor parent child communication, and delinquent peers (Fite, Wynn, & Pardini, 2009). Neighbourhood and family social conditions were associated with racial/ethnic disparities in rates of violence for White, Black and Mexican youth in a large American city (Sampson, Morenoff, & Raudenbush, 2005). Schwalbe (2008) makes the point that important differences in risk profiles can be masked by aggregate rather than subgroup analsyes.

A related reason that diverse subgroups should be examined is that factors such as gender, race, ethnicity, socio-economic status and age are implicated in the everyday practices of juvenile justice. Thus, the offences that young people are charged with, the diversionary options available to them, the convictions they attract, and the interventions offered are known to be influenced by such socially constructed identifiers (Cunneen & White, 2007; McGrath, (in press); Rodriguez, 2010; Tonry & Melewski, 2008). These factors influence key aspects of risk assessment such as criminal history, detention, and likelihood of recidivism. The direct and indirect effects of such factors may result in what Rodriguez (p. 393) refers to as “cumulative disadvantage” within the juvenile justice system.

A third reason that risk-need assessment should be examined at the subgroup level is to clarify for users how such identities (e.g., gender, race/ethnicity) can be taken into account and whether results from a risk assessment inventory are collectively appropriate. The standards for educational and psychological testing (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1999) make it clear that fairness in testing is achieved, in part, by lack of bias in test score performance for identifiable subgroups. Such assurances are also critical in the psycho-legal arena where the standards of evidence for psychological tests are becoming

more rigorous (Medoff, 2003; Tolman & Mullendore, 2003). Certainly risk-need assessment has it critics (Baird, 2009; Hart, Michie, & Cooke, 2007) and decisions may be challenged for not giving due attention to how representative the normative group is for the indivdual being assessed (Canadian Association of Elizabeth Fry Societies, 2008; Hanson & Howard, 2010). Recently, norms have been provided by gender and setting (community versus custody) for the YLS/CMI 2.0 (Hoge & Andrews, 2011). These norms ground the percentile rank of an individual's score to one of four subgroup and thereby indicate risk-need relative to males or females either in the community or in custody. However, the resulting risk/need classifications have not been compared for predicting continued criminal activity. Based on existing research, there is little doubt that a predicitve association exists, but the re-offending rates by subgroups have not been reported. Analysis of subgroup differences at the item and domain levels, as well as investigating the link between classification and recidivism, will contribute to the psychometric and practical benefits of risk-need inventories for diverse juvenile populations .

We stated that the existing literature on subgroup difference with the YLS inventories is limited. Few studies provide item level analyses and those that do often report only aggregate results (Onifade, Davidson, Livsey, et al., 2008; Thompson & Pope, 2005). However, Flores, Travis and Latessa (2003) did provide item analyses for gender and racial (White, Non-White) subgroups in a sample of aproximately 1500 Ohio youth under custodial and probation supervision. Subgroup differences in item endorsement proportions were common. On the basis of gender, there were significant differences on just over half of the items; whereas by race, two thirds of the items differed. Such item analyses are particularly helpful as they reveal the source of differences in the broad risk domains. For example, the significantly higher score for females on the family risk domain was due to females being more likely to experience inappropriate discipline and poor relations with parents, especially

mothers. For non-Whites, a higher incidence of occasional and chronic drug use augmented their overall score on the substance abuse domain. Research with other risk inventories also indicates that gender and racial/ethnic differences are likely to occur at the item level (Baker, Jones, Roberts, & Merrington, 2003; Schwalbe, Fraser, & Day, 2007).

Other studies report comparisons of the total inventory scores for subgroups. The user's manual for the YLS/CMI (Hoge & Andrews, 2006) provides normative data for 173 male and 91female juvenile offenders. This data came from Jung and Rawana (1999) who reported that there were no significant differences by gender on the total inventory score. The YLS/CMI 2.0 manual (Hoge & Andrews, 2011) reports no gender difference in total score based on much larger samples. Flores et al. (2003) found that females scored 1.3 points higher than males overall. Schmidt, Hoge and Gomes (2005) reported no significant mean differences on the total inventory score for 67 male and 40 female offenders referred for mental health assessments. Nevertheless, females were approximately 3 points higher on the total score. In a Scottish study of youth in secure and residential settings, males (N= 52) scored approximately 7 points higher than females (N=42) on the YLS/CMI total score (Marshall, Egan, English, & Jones, 2006). Onifade, Davidson, Campbell, et al. (2008) found that girls on probation (N=90) scored a significant 2.6 points lower than boys on probation (N= 238) in a midwest sample from the United States. The total inventory score was also reported as significantly higher for Native compared with non-Native youth (Jung & Rawana, 1999) and for non-Whites compared with Whites (Flores et al., 2003)). New normative data for the YLS/CMI 2.0 show effect sizes bordering on small for total score difference between white and non-white offenders in custody (d = 0.19) and in the community (d = 0.18).

As might be expected, total score differences often emerge from the risk inventory domains. Differences across subgroups were common in the YLS study of Flores et al. (2003). On the basis of gender, four domains differed and five domains differed in the racial

analyses. In their Canadian sample of offenders, Jung and Rawana (1999) found that Native youth scored higher on negative peer relations, substance use, and engagement with recreational actvities. Schmidt, Campbell, and Houlding (2011) reported that scores for females were significantly higher than for males on the family subscale. Hoge and Andrews (2011) reported minimal domain differences by gender and race/ethnicity. A few differences were evident in their custodial samples. For example, females had more family problems than males and non-white males had more negative peer relations than white males. Collectively, these studies show that subscale and total inventory scores on YLS inventories do differ by subgroups. However, there is considerable variability in the direction and extent of differences depending on subgroup charcteristic, sample size, setting and jurisdition.

The predictive validity of juvenile risk assessment instruments in regard to gender has been examined in a number of studies. Schwalbe (2008) identified 19 studies that used various second and third generation risk assessment instruments and provided predictive validity data by gender. These studies from 1999 to 2007 produced 49 effect sizes (25 for males, 24 for females). In a meta-analysis of unweighted point-biserial correlation coefficients, Schwalbe found and average effect size of $r = .26$ for males and $r = .27$ for females. Only four of the studies in Schwalbe's meta-analysis used the Youth Level of Service/Case Management Inventory for which the predictive validity indices were $r = .32$ and $r = .40$ for males and females respectively. Schwalbe concluded that the design of most contemporary risk assessment inventories "leads to risk classifications with similar levels of predictive validity for male and female offenders" (p.1377). Olver et al. (2009) also present meta-analytic results for juvenile risk measures using a different but overlapping sample of studies to Schwalbe (2007, 2008). A gender breakdown of nine predictive validity studies that used youth versions of the Level of Service Inventory (LSI) showed that the mean weighted effect size for general recidivism was $r_w = .33$ for males and $r_w = .36$ for females.

The predictive validity of juvenile risk assessment by race/ethnicity is more difficult to aggregate across studies due to both country and even jurisdictional differences in the diverse subgroups represented. In a subset of five of studies, Olver et al. (2009) found r_w = .35 and r_w = .36 respectively for Aboriginal Canadian youth compared with Non-Aboriginal Canadian youth. The authors concluded that the predictive validity of youth adaptations of the LSI was supported for Aboriginal youth (p. 347). Bechtel, Lowenkamp, and Latessa (2007) compared the predictive validity of the YLS/CMI for White and non-White (largely African American) youth in a large sample in the State of Ohio. They found significant, but not statistically different, correlation coefficients of r = .21 and r = .16 respectively. Predictive validity by race/ethnicity has been found to vary with other juvenile risk instruments in other jurisdictions. In two studies (Schwalbe, Fraser, Day, & Arnold, 2004; Schwalbe, Fraser, Day, & Cooley, 2006), differences were found between White and African American youth in the predictive validity of the nine item North Carolina Assessment of Risk (NCAR). In a subsequent study (Schwalbe, Fraser, & Day, 2007), the NCAR and an expanded version of the instrument produced somewhat different AUC values for Non-Latino White versus African American juveniles. However, the 95% confidence interval for those indices all overlapped.

In summary, the literature shows that the psychometric properties of risk-need assessment instruments for juvenile offenders may or may not be similar across some of the distinguishable subgroups with which they are used. For both fairness and utility, it is important to investigate such differences. This needs to be done relative to the risk instrument and the juvenile justice jurisdiction in which the instrument is used. Moreover, it is important to elucidate whether any psychometric differences that do emerge, even if they are relatively small, impact on the way the inventory is used and decisions that are based on the inventory. This is seldom considered in the research literature. For example, the equivalence of

predictive validity indices for subgroups in extant research has lead to reassurances that one size of juvenile assessment inventory fits all. In practice though, youth are classified into several categories of risk-need that are the basis for different courses of action in keeping with risk/need/responsivity principles (Andrews, Bonta, & Hoge, 1990; Hoge & Andrews, 2011). The relationship between such categories and external criteria is typically neither elucidated nor compared for subgroups. For practitioners and policy makers, validity coefficients and non-gendered classifications do not answer the question 'how likely to re-offend is a female assessed as moderate risk?'. Some have taken the position that risk-need inventories should be adapted or developed for particular subgroups. Emeka and Sorenson (2009) argue for gendered risk assessment instruments. Empirically they showed how the predictive accuracy of a second generation risk measure varied when non-gendered risk categories were applied to females. Two year recidivism rates associated with the same risk categories differed up to approximately 11% for males versus females. It is this kind of extended analysis that should be undertaken to complement the utility of risk inventories and to inform the debate on crime and subgroup dynamics.

The research reported here examined both gender and ethnic subgroups in a large sample of juvenile offenders in New South Wales, Australia. In this jurisdiction, the Australian Adaptation of the Youth Level of Service/Case Management Inventory (YLS/CMI-AA; Hoge & Andrews, 1995) has been used to assess risk and needs. Data from a large sample allowed: 1) examination of the item, domain and total risk scores by gender and ethnicity, and 2) exploration of predictive validity and predictve accuracy for the same subgroups.

Method

Data

Approval to access the data for this study was granted by the appropriate university and government ethics committees. Data were provided by the New South Wales (NSW) Department of Human Services (Juvenile Justice) where the YLS/CMI-AA was adopted in October 2002 as a significant part of the assessment policy and procedure. Information on the adaptation process and the preliminary psychometric data were provided by Thompson and Pope (2005). Like the parent version, the instrument includes risk/need items over eight domains: (1) *prior and current offences*; (2) *family and living circumstances*; (3) *education/ employment*; (4) *peer relations*; (5) *substance abuse*; (6) *leisure/recreation*; (7) *personality/ behavior*; (8) *attitudes/ beliefs*. There are 47 items in total on the inventory with all but one scored in a binary fashion to indicate whether the operationally defined risk factor applies to the young person. One item related to age at first court order is scored 0, 1, 2 with more weight given to younger offenders. Domain subscores vary between 3 and 9 while the total score can range from 0 to 48.

In a computerized format, the inventory was completed for all young persons under various forms of supervision in the community and some young people in custody. The inventory was completed by Juvenile Justice Officers who receive training on its use and are required to develop a case management plan based on the risk/need findings. The data set included scores for 6890 inventories completed up to early December 2005. The data were screened to remove incomplete, duplicated and mistaken entries. The data were then restructured to link follow-up YLS/CMI-AA assessments of the same individual. Departmental policy required repeated completion of the inventory at regular intervals. The result was 6632 inventories relating to 4238 juvenile offenders from 40 juvenile justice locations throughout the state. We were able to obtain recidivism data for 4138 of these individuals From these, we eliminated four cases with data errors, 296 cases from 2002 when the inventory was being rolled out, and 270 cases where the inventory was used to assess

youth serving custodial sentences. The final sample on which the results are based therefore incorporates all first administered YLS/CMI-AA inventory results for 3568 young people that were completed between January 02, 2003 and December 02, 2005.

Age at the time the inventory was completed was distributed as 16.8% under 15 years, 41.3% aged 15 and 16 years, and 42.0% aged 17 years and over. The sample was 15.7% female and 84.3% male. There were proportionately more males (43.4%) than females (34.2%) aged 17 years and over. Mean age for males (M = 16.51, SD = 1.50) was significantly higher than for females (M = 16.30, SD = 1.39; $t(3566)$ = -3.01, p = .003) Breakdown by ethnic background was based on the self-report of the young person and family to standard assessment questions. *Australian Indigenous* status (29.5%) included youth of Australian Aboriginal and/or Torres Strait Islanders origin as defined by the Australian Bureau of Statistics. Youth who were neither of Indigenous nor of other ethnic background, were categorized as *Australian* (44.3%). Youth having a non-Australian cultural background (88 countries/ethnic groups represented) were categorized as *Australian Ethnic* (21.4%). The remaining cases were instances of missing or unknown information (4.7%). There were proportionately more *Australian Indigenous* youth under 15 years (24.8%) than *Australian* (15.2%) and *Australian Ethnic* (11.0%) youth. Overall, Indigenous youth were younger (M = 16.11, SD = 1.66) than youth in the other two groups (respectively M = 16.57, SD = 1.43 and M = 16.72, SD = 1.27; $F(2, 3397) = 45.07$, $p < .001$).

Recidivism was defined as a re-offence occurring within one year of the administration of the YLS/CMI-AA that resulted in a conviction. Where the young person was charged and convicted of more than one offence, the offence attracting the most severe penalty was selected as the index offence. The analyses were based on the date of the actual offence rather than the date of conviction, which varied depending on the police and court processing times. This calculation has the advantage of representing “real time” to offending.

It reduced extraneous sources of time variability in our examination of the predictive reach of the inventory. The re-offending data were retrieved from the New South Wales Bureau of Crime Statistics and Research reoffending database (ROD). ROD contains all finalised juvenile and adult criminal appearances in NSW higher courts (District and Supreme), NSW Local Courts and the NSW Children's Court (Hua & Fitzgerald, 2006). There was ample time (approximately 11 months) for a conviction to be finalized and registered beyond the "real time" follow-up period of one year. Within 3 months of the first YLS, 20.9% of the sample had re-offended. Within 6 months, 33.7% had re-offended and within 12 months, 50.7% of the sample had re-offended.

Results

Item Analysis

First, endorsement proportions were calculated for all items and comparisons made by gender using Pearson's chi-square statistic, Cramer's V, and a conservative probability level of .001. The Cramer and Pearson statistics are both measures of association. However, Cramer's V ranges between 0 and 1 and thereby gives both a measure of significance and effect size similar to a correlation coefficient (Field, 2009). There were significant gender differences for a number of items. Males (26.2%) were more likely than females (19.3%) to have three or more prior offences. Females (57.2%) were more likely than males (46.9%) to use alcohol only occasionally, although equal proportions of both genders (20%) used alcohol regularly. Females were more likely to not participate in organised leisure or recreational activities (70.7% versus 58.7 % for males), more likely to display tantrums (38.1% versus 27.5%), and more likely to be verbally aggressive (42.4% versus 31.3%). The domain that showed the most consistent pattern of gender differences though was *Family and Living Circumstances*. Five of the seven items were significantly different ($\chi^2(1)$ from

12.19 to 40.05, p <.001, Cramer's V from .058 to .106, p < .001). Females were more likely to experience inadequate monitoring, lack of parental control, inappropriate discipline, poor relations with mother/stepmother and homelessness. The endorsement of these items was typically 7-10% higher for females with the biggest differential for difficulties with mother (42.4% for females versus 28.9% for males).

The same analytical approach was used to examine item endorsement by the three ethnic subgroups. All but five of the 47 items showed significant differences and all risk domains were represented ($\chi^2(2)$ from 14.80 to 289.98, p <.001, Cramer's V from .066 to .292, p < .001). For just over half of these differentially endorsed items, the pattern was one of more Indigenous youth represented than Australian youth who were in turn represented more than Ethnic youth. This pattern varied in degree. For example, the family domain item for anti-social values in the family was endorsed for 47.3% of the Indigenous youth, 22.2% of Australian youth and 14.4% of Ethnic youth. The proportion of youth not seeking help was 36.6%, 31.2% and 27.1% for the same subgroups respectively. Although this pattern was dominant, several others were also evident but only one prominent. For approximately one quarter of the item differences, the Indigenous and Australian subgroups were similar but the Ethnic subgroup was less frequently represented. All four of the differentially endorsed items in the Personality/ Behavior domain showed this pattern. For example, verbal aggressiveness was reported for 36.8% of Indigenous youth, 36.5% of Australian youth and 22.6% of Ethnic youth.

Domain Analyses

The mean and standard deviation of domain scores by subgroups are presented in Table 1. Parametric statistical tests were used to compare domain score by subgroups. Subgroup distributions of domain scores were predominantly positively

skewed and platykurtic. However, large subgroup samples make it likely that the assumption of normality for sampling distributions was satisfied (Field, 2009, p.134). Whenever homogeneity of variance assumptions were violated, adjusted *t* and *F* test values were used. Conservative *p* values were chosen to protect familywise error rate due to multiple tests. Post hoc ANOVA comparisons used the Games-Howell procedure due to differing subsample sizes and heterogeneity of subgroup variances (Field, p. 374).

By gender, four of the domain scores differed significantly (*family and living circumstances; substance abuse, leisure/recreation, personality/behavior*). The largest gender difference was just over a half point higher for females on the family domain ($t(3566) = 6.59$, $p < .001$) and represented a small effect size ($r = .11$). By ethnic subgroups, all domain scores varied significantly (Welch's $F(2, 1900$ approximate) from 21.11 to 93.3, $p < .001$). The effect size for these differences ranged from $r = .11$ for *education/employment* to $r = .23$ for *prior and current offences*. Post hoc pairwise comparisons between subgroups in each domain showed that on five domains (*prior and current offences, family and living circumstances, education/employment, leisure/recreation, attitudes/ beliefs*) the Indigenous group scored significantly higher than the Australian group who scored significantly higher than the Ethnic group ($p \leq .016$). The mean score difference between groups varied across domains as can be seen in Table 1 and amounted at most to approximately 1 point. On the remaining domains (*peer relations, substance abuse, personality/ behavior*), the Indigenous group was always significantly higher than the Ethnic subgroup ($p < .001$) with the Australian subgroup in-between and not significantly different from either one group or the other.

The cumulative effect of domain differences by subgroups is represented in the total YLS/CMI-AA risk/need score and this was significant by both gender ($t(3566) =$

3.07, $p < .01$) and ethnic (Welch's F (2, 1935) = 97.55, $p < .001$) categorisations. The gender difference in mean total score was only 1.4 points higher for females and represents a very small ($r = .05$) effect. The overall effect size for ethnic subgroups was between small and medium ($r = .22$). Post hoc comparisons showed that the pairwise differences between all ethnic subgroups were significant ($p < .001$). The Indigenous mean total score was 2.6 points higher than the Australian mean which in turn was 3.4 points higher than the Ethnic mean. The 6 points differential in total risk score between the Indigenous group and the Ethnic group was a moderate effect size ($r = .31$).

Predictive Validity

The relationship between the YLS/CM-AA total score and the one year dichotomous recidivism measure was computed for the sample as a whole and also separately for subgroups. Both the point biserial correlation and receiver operating characteristic (ROC) curve analysis were used. ROC analysis plots the true positive rate (sensitivity) against the false alarm rate (1 minus specificity) for all possible cutting values of the risk measure. The resulting area under the curve (AUC) is a useful and commonly reported index of predictive accuracy (Bennett et al., 1999; Rice & Harris, 1995). The results are presented in Table 2. The correlation coefficients were between $r = .17$ for the Indigenous subgroup and $r = .27$ for males. The validity coefficient for males was not significantly different from that of females. However, the coefficient for the Indigenous group was close to being different from that of the other two ethnic groups ($p < .06$). The validity coefficient for the total sample was $r = .26$ which is a medium effect size for a point-biserial correlation with a re-offending base rate of 50% (Rice & Harris, 2005). AUC values were between .604 for the indigenous subgroup and .659 for males. The 95% confidence intervals for subgroup *AUC*s

overlap, although those for the ethnic subgroups show a pattern of near difference similar to that for the correlation coefficients. Subgroup confidence intervals also overlapped with the confidence interval for the combined sample AUC = .652. The latter value indicates that there was a 65.2% probability that a randomly selected recidivist would score higher on the YLS/CMI-AA total score than a randomly selected non-recidivist.

Predictive Accuracy

To show how categories of risk based on the overall sample impact on prediction accuracy for gender and ethnic subgroups, cut scores were identified. The cut score ranges were not exactly the same numerically as those of the YLS/CMI (Hoge & Andrews, 2006) and YLS/CMI2.0 (Hoge & Andrews, 2011) because the Australian adaptation has a higher maximum score. However, the cut scores we derived were for proportions of our sample equivalent to those categorised by the cut scores of the parent version. Specifically, the *low risk* category was for a total score between 0 and 15 which accounted for the 45.6% of our overall sample. The *moderate risk* range (16-30) accounted for 43.4% of the sample. The *high* and *very high* categories were combined to included total scores between 31 and 48 which represented 11% of the sample. These categories were combined because in the parent version, the *very high* category is for the top 1% of scores. Even with our large sample, the actual number of scores in this range is too few to permit meaningful analysis of predictive accuracy. Bechtel et al. (2007) also combined the *high* and *very high* YLS categories.

Table 3 shows the rate of re-offending after 1 year for the young people in these risk categories. The results are for the total sample and for gender and ethnic subgroups. To begin, it is of interest to compare the rate of re-offending for the

combined male and female sample with the actual rate of re-offending for males and females separately. For example, using the total sample, one would conclude that of low risk male and female juvenile offenders, 38.4% will re-offend within one year. However, for females with risk scores in that range, 29.4% re-offend while for males the rate is 39.8%. Similar gender disparities are evident in the *moderate* and *high risk* categories. The gender differences in re-offending rate within each risk classification were significantly different (*low* ($\chi^2(1) = 8.29$, $p = 004$, Cramer's V = .073, $p = .003$; *moderate* ($\chi^2(1) = 24.09$, $p < .001$, Cramer's V = .126, $p < .001$; *high* ($\chi^2(1) = 6.24$, $p = 012$, Cramer's V = .133 $p = .008$). Looking at the ethnic subgroup data in Table 3, one can also compare re-offending based on the combined sample data with the rates for subgroups. For example, 71.9% of all young people in the *high* risk category re-offend. However, for that risk category, the actual rates are 75.5% for Indigenous youth, 69.6% for Australian youth, and 68.3% for youth with other ethnic backgrounds. The disparities are more marked between ethnic subgroups for youth in the *low* and *moderate* risk categories. The ethnic differences in re-offending rate were significant for youth in the *low* risk category (($\chi^2(2) = 39.92$, $p < .001$, Cramer's V = .162, $p < .001$) and *moderate* risk category (($\chi^2(2) = 21.63$, $p < .001$, Cramer's V = .120, $p < .001$), but not for youth classified as *high* risk. The bottom row of Table 3 shows the base rate of re-offending for all subgroups. Females and ethnic youth have the lowest rate around 40%, while males and Indigenous youth are higher at 53% and 63% respectively.

The coordinate points of a ROC curve also allow estimates of the hit and false positive rates associated with scores on the inventory. The ROC values are for the average of two consecutive observed values of the total risk score. They show the proportion of re-offenders correctly identified by using the score as a cut point and the

proportion misclassified at that score. For our total sample, a prediction that all juveniles with scores above 19.5 re-offend was associated with a hit rate of 51% (the overall base rate) and a false positive rate of 28%. The subgroup ROC curves showed differences by gender and ethnicity. The same cut score of 19.5 was associated with a hit rate of 50 % for males and 57% for females, but with false positive rates of 27% and 35% respectively. By ethnicity, and using the same cut score, the hit rates were 58% (Indigenous), 51% (Australian) and 38% (Ethnic) with associated false positive rates of 41%, 30% and 19% respectively.

Discussion

The results of this study showed both gender and ethnic differences for a large sample of community based juvenile offenders who were evaluated using the Australian Adaptation of the Youth Level of Service/Case Management Inventory. Differences were found at the item level, across domain scores and on the total inventory score, but not for predictive validity indices. However, predictive accuracy rates using either risk categories or coordinates of the ROC curve did show differences by gender and ethnicity.

Differences in item endorsement proportions were observed for 21% of the items by gender and 89% by ethnicity. This is consistent with the high frequency of item differences for subgroups found by Flores et al. (2003). Given the large number of items on the YLS inventory, differences should be expected when various subgroups are compared. This is because the content covered in most domains is sufficiently wide to incorporate correlates of offending that may distinguish some groups more than others. Schwalbe, Fraser and Day (2007) note this as an advantage of longer over shorter risk inventories. With respect to gender, it is increasing well established that there are precursors and pathways to crime that characterise young females more so than males. Family problems, maltreatment and abuse are frequently mentioned as propelling girls and young women into a range of risky and anti-

social behaviors (Belknap & Holsinger, 2006; Emeka & Sorensen, 2009; Odgers, Robins, & Russel, 2010). The current study showed that content breadth in the family and living circumstances domain is sufficient to reflect some aspects of this pattern. Females, more than males, had troubled family interactions including lack of control, inappropriate discipline, poor relationship with mother, and homelessness. Other item differences were not as concentrated in any single domain . Nevertheless, some of those differences do echo findings about female compared to male juvenile offending. For example, girls were more likely to display tantrums and more likely to be verbally aggressive. Aggression and violence by adolescent females can be qualitatively different from that of males. Females are less likely to be involved in serious forms of violence and adolescent girls are likely to engage in relational acts of aggression against the adults and peers in their life (Conway, 2005; Odgers, Moretti, & Repucci, 2005).

Item endorsement differences across the three ethnic subgroups of this study also reflect circumstances that have been noted as differentially relevant to juvenile offending. Almost all of the items differed by ethnicity. For this reason, interpretation is best conveyed at the domain level. The Indigenous subgroup was higher than other two groups on five domains (*prior and current offences, family and living circumstances, education/employment, leisure/recreation, attitudes/ beliefs*). The health, social, and economic inequities of Indigenous Australians is well documented (Australian Institute of Health and Welfare, 2011b), as is their alarming over-representation in the criminal justice system (Australian Institute of Health and Welfare, 2011a). Thus, it is not surprising that the Indigenous youth in this study scored higher on items related to previous and current offending. This was the largest effect size at the domain level. Some explain such findings as a manifestation of systemic biases that have a sordid history and continue through the contemporary criminalisation of Indigenous people (Commonwealth of Australia, 2011; Cunneen & White,

2007). An alternative view is that certain forms of criminality are more prevalent among Indigenous youth due to colonial history, social and economic problems, and cultural traditions (Blagg, 2008; Commonwealth of Australia, 2011; Weatherburn, Snowball, & Hunter, 2008).

The higher scores of Indigenous youth on items and domains concerning family circumstances, education/employment and leisure/recreation reflect the co-occuring risks that are prevalent for Indigenous Australians. On numerous determinants of health and welfare, Indigenous Australians rate substantially worse than non-Indigenous Australians (Australian Institute of Health and Welfare, 2011b). The current study also showed that Indigenous youth scored highest on the attitudes/belief domain. They recorded proportionately higher item frequencies for four of the five items including having anti-social values and defying authority. Consistent with this, *anti-social values in family* (an item in the family risk domain) was endorsed for 47.3% of the Indigenous youth, 22.2% of Australian youth and 14.4% of Ethnic. Anti-social values and attitudes are a key ingredient of criminality (Andrews, Bonta, & Wormith, 2006; Emeka & Sorensen, 2009). Nevertheless, prejudice can lead to negative stereotypes about members of marginalised groups. This is especially the case for Indigenous Australians who suffer many forms of discrimination (Cunneen & White, 2007; Pedersen & Barlow, 2008). Cunneen and White (2007) also suggest that some aspects of Aboriginal youth offending can be regarded as defiance and resistance toward dominant non-Indigenous institutions.

Item and domain analyses showed that risk factors were less prevalent for the group referred to as *Australian Ethnic.* We derived this homogenous category from information collected by the juvenile justice department rather than a priori expectations about specific ethnic risk differences. *Australian Ethnic* youth were those identified as having a non-Australian cultural background. The ethnic backgrounds most reported were Lebanese, New

Zealander and Vietnamese, and each comprised 9-10% of the larger *Australian Ethnic* group. However, Pacific Islanders, including youth identifying as Polynesian, Maori, Samoan, Tongan, Fijian and Cook Islanders, collectively made up 28% of the subgroup. Informed analysis shows a variety of interacting cultural, language, social, economic and global factors at play for these groups in Australia. However, inaccurate metaphors and misleading media reports can reduce this complexity to moral panic over criminal proclivities (Cunneen and White, 2007; Dodge, 2008). The reduced prevalence of risk factors for the *Australian Ethnic* subgroup as a whole suggests otherwise. The findings also show how varied the ethnic mix can be in any jurisdiction. Racial, ethnic and cultural categorizations are poor proxies for the complex relational issues that underpin them (Markus, 2008; Phinney, 1996).

The item and domain differences by subgroups led to differences in the total risk/need score on the inventory. These differences amounted to females scoring 1.4 points higher than males. Indigenous youth scored 2.6 points higher than Australian youth and 6 points higher than the Ethnic youth. Although small in statistical effect size, some might consider such differences indicative of test bias. However, mean differences between groups do not alone constitute evidence that a psychological measure is biased. If differences are consistent with what is known about offenders from specific subgroups, such findings confirm that domain relevant content is represented in the inventory and it is sensitive to those differences.

The predictive validity results of the current study were in keeping with those reported elsewhere. Validity indices (point biserial *r* and *AUC*) for the total sample were similar to those for the YLS and other second and third generation risk assessment instruments as reported in recent meta-analyses (Olver et al. 2009; Schwalbe, 2007). By gender, the point biserial coefficients in our sample were only marginally smaller than the mean weighted effect sizes for males and females from

nine previous YLS studies (Olver et al. 2009). The predictive validity effect sizes by gender and ethnic subgroups reported in Table 2 did not differ significantly. However, for the Indigenous group, the correlation coefficient and the *AUC* confidence interval border on a lower degree of magnitude. Nevertheless, the results provide further support for the YLS inventories and indicate that predictive validity is quite robust across a number of subgroups.

The predictive accuracy results in Table 3 relate to risk-need classification and to issues of test utility in ways that are not self-evident from predictive validity indices. Using risk-need categories equivalent to those suggested in the YLS/CMI manuals (Hoge and Andrews, 2006, 2011), the results showed that 1 year re-offending rates increase substantially across three classification levels. The accuracy rates for those same categories applied to gender and ethnic subgroups showed some variation and more than might be expected from the predictive validity indices. This is because subgroup accuracy rates reflect a combination of effects beyond small differences in validity coefficients. These include differences in the subgroup distribution of scores and mean total score, differences across subgroups in the base rate of re-offending, and the effects of continuous raw scores being converted to discrete raw score categories. Moreover, the results in Table 3 show how classification accuracy rates based on the total sample are tipped toward the majority subgroup (e.g., males or Australian) due to their proportional over-representation.

Predictive accuracy data are important for two reasons. First, they provide risk inventory users and policy makers with easy to understand information. Gendreau and Smith (2007) have made the point that plain language statistics are most helpful in explaining prediction and treatment effects in offender research. To that end, percentage differences in recidivism are basic units of interest. Gendreau and Smith

note, “Nothing could be simpler and more readily understandable for all concerned” (p.1541). Predictive accuracy data are also important because they extend our understanding of test performance beyond that provided by predictive validity indices. Prediction rests on more that the correlation between the test and the criterion. Equally important is the method that is used to relate test scores to the criterion and whether that leads to different predictive outcomes for different groups (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1999, Test Standard 7.6; Betts et al., 2008). The current study showed that when categories of risk are linked to 1-year re-offending, there are varying subgroup differences in accuracy. That pattern was also evident in the results drawn from coordinates of the subgroup ROC curves.

Implications

The limited representation of item content for various subgroups may be viewed as a potential source of bias on inventories such as the YLS (Emeka & Sorensen, 2009). However, we think customising item content is not necessary. The current study showed that the YLS/CMI-AA items have a degree of sensitivity to gender and ethnicity. Although offending pathways and dynamics vary across prominent subgroups, common continua of risks and needs are evident with score distributions overlapping (Gavazzi et al., 2006; Hubbard & Pratt, 2002). Also, risk factors are not independent but co-occurring (Dixon, Howie, & Starling, 2005; Farmer & Farmer, 2001). Not all correlated risk variables need to be psychometrically represented to achieve the desired objectives. On that account, the quantifiable part of YLS inventory (Parts I and II) should not be confused with the full assessment of an adolescent offender. The needs and special considerations often noted as being overlooked for girls and ethnic subgroups are listed as important matters for evaluation

(Part III of the inventory and Appendix B of the user's manual; Hoge & Andrews, 2006, 2011). We believe contextualised and idiographic considerations are best made at this level and factored into the overall assessment along with the psychometric results.

We also believe that it may not be necessary to develop separate subgroup norms for YLS inventories. A step in the direction of subgroup norms has been taken with the YLS/CMI 2.0 (Hoge & Andrews, 2011) where new cut points are provided for males and females in custodial and community settings. However, those norms only ensure that the same proportion of offenders occurs in the same risk-need categories. The predictive accuracy of the resulting classifications has not been examined and may not be an improvement on using a common set of norms. Transparency about predictive accuracy rates for subgroups if common norms are used is important as the results of our study show. Within group norming and other forms of test score adjustment is complex (Kehoe & Tenopyr, 1994). It has proven contentious and even unlawful when cognitive ability tests have been used for employment selection (Sackett & Wilk, 1994; Schmitt, Sackett, & Ellingson, 2002). It is further compounded by the inadequacy of identifiers based on sex, race, ethnicity, culture, and national origin to reflect common underlying differences. These are indicators of surface-level diversity and may not map well onto deep-level diversity (Eagly & Chin, 2010). One must also decide which ethnic subgroups warrant separate norms and which do not.

We believe that inventories such as the YLS/CMI-AA can be used fairly with a diverse offender population. Test fairness involves considering the outcomes of testing in addition to the technical properties of tests (American Educational Research Association, American Psychological Association, National Council on Measurement

in Education, 1999). Test fairness involves policy decisions about how tests are used and the ways that performance characteristics such as differential prediction may be dealt with (Sackett & Wilk, 1994). Considering the results of the current study, we think the YLS/CMI-AA can be used fairly with a common set of classification norms under the following conditions. First, several risk categories should be determined that are associated with increasing levels of risk. Among the various considerations for optimal norms, cut scores should be chosen to minimise differential prediction for dominant subgroups (McGrath & Thompson, 2011). The differentials to be minimised are those between the common norms and subgroups, not those between subgroups per se. Second, the resulting risk categories should be communicated via descriptors such as *low, moderate, high* and *very high* risk of re-offending. This is essentially a crude form of banding of test scores (Sackett & Wilk, 1994). It implies that all youth are similar within each band on the construct of interest (i.e., recidivism risk). In the current study, score band descriptors *low*, *moderate* and *high* were faithful to the increasing relative risk for all of the subgroups. Third, it is important to consider whether common norms are justified for the types of decisions that are being made. This is the *evaluative and prescriptive* function of risk assessment (Schopp, 1996). Those making decisions need to judge whether differential action based on different categorizations is justified. Some decisions may well tolerate the kind of variations that we found in predictive accuracy rates across groups. Other decisions may warrant a degree of override or the weighing of social values and priorities. It is essential that those responsible for such decisions have specific information about differential prediction and the limitations of the related social categorisations. This is consistent with contemporary legal and ethical standards (Bazelon, 1982; Medoff, 2003; Tolman & Rotzien, 2007). Fourth, users of the inventory need to be trained and maintain

currency with its relevance for the populations served. Research related to juvenile offending for gender and racial/ethnic subgroups should be central to that training. Users need to properly understand norms and they need to know the margins of error associated with norms. It is especially important of take into account the needs and special considerations that are supplementary to the quantifiable primary risk factors.

Limitations

A number of limitations qualify the results and interpretations of the current study. We only examined differences in item endorsement by subgroups. More sophisticated analysis such as differential item functioning (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1999; Sackett, Schmitt, Ellingson, & Kabin, 2001) or item response theory models (Embretson, 1996) are worth pursuing. Our criterion measure was derived from official conviction data. However, the follow-up period of 12 months was based on the earlier date of actual offending not the subsequent court finalization date. This is in keeping with the real time case management intent of risk assessment, but a longer follow-up period would contribute to understanding short versus long term recidivism. Variation in the follow-up period and in the operational definition of re-offending can account for differences in predictive validity across studies (Singh & Fazel, 2010). Bias in the criterion measure can be a source of differential prediction (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1999; Sackett & Wilk, 1994). The conviction process could be vulnerable to biases for the subgroups we examined. Another limitation is that we examined gender and ethnic group differences independently. These and other social identifiers intersect and much is to be gained from examining the interaction of such “social locations” (Bechtel et al. 2007; Burgess-Proctor, 2006; Rodriguez, 2010). However, the research trends in this direction strengthen further our reasons for

eschewing separate inventories and separate norms. Increasing differentiation yields more subgroup permutations that could conceivably lead to a progression of inventory customisations or subgroup norms. Finally, the position we elaborate on the use of common norms relates to our particular sample of community offenders. We think this advances thinking on timely and important issues but recognise that jurisdictional and setting difference may warrant a different argument. For example, the context and circumstances of youth in custody might justify both modifications to risk-need assessment approaches as well as specific norms. In conclusion, subgroup analysis of risk measures for youthful offenders does advance our understanding of how they may be used to serve juvenile offenders and juvenile justice.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior, 17*, 19-52.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency, 52,* 7-27. doi:10.1177/0011128705281756.

Australian Institute of Health and Welfare. (2011a). *Juvenile justice in Australia 2008–09. Juvenile justice series no. 7*. Cat. no. JUV 7. Canberra: Author. Retrieved from http: //www.aihw.gov.au/publication-detail/?id=10737418606

Australian Institute of Health and Welfare. (2011b). *The health and welfare of Australia's Aboriginal and Torres Strait Islander people, an overview 2011*. Cat. no. IHW 42. Canberra: Author. Retrieved from http://www.aihw.gov.au/publication-detail/?id=10737418989&tab=2

Austin, D. (2010). Narratives of power: Historical mythologies in contemporary Quebec and Canada. *Race & Class, 52*(1), 19-32. doi:10.1177/0306396810371759

Baker, K., Jones, S., Roberts, C., & Merrington, S. (2003). *ASSET-The evaluation of the validity and reliability of the Youth Justice Board's assessment for young offenders. Findings from the first two years of the use of ASSET*. Oxford, United Kingdom: Probation Studies Unit, Centre for Criminological Research, University of Oxford.

Baird, C. (2009). *A question of evidence: A critique of risk assessment models used in the justice system.* Madison, WI: National Council on Crime and Delinquency.

Bazelon, D. L. (1982). Veils, values, and responsibility. *American Psychologist, 37*, 115-121.

Bechtel, K., Lowenkamp, C. T., & Latessa, E. (2007). Assessing the risk of re-offending for juvenile offenders using the Youth Level of Service/Case Managment Inventory. *Journal of Offender Rehabilitation, 45(3/4*), 85-108. doi:10.1300/J0t6v45n03_04

Belknap, J., & Holsinger, K. (2006). The gendered nature of risk factors for delinquency. *Feminist Criminology, 1,* 48-71. doi:10.1177/1557085105282897.

Bennett, K. J., Lipman, E. L., Brown, S., Racine, Y., Boyle, M. H., & Offord, D. R. (1999). Predicting conduct problems: Can high-risk children be identified in kindergarten and grade 1? *Journal of Consulting and Clinical Psychology, 67*, 470-480.

Betts, J., Reschly, A., Pickart, M., Heistad, D., Sheran, C., Marston, D. (2008). An examination of predicitve bias for second grade reading outcomes from measures of eraly literacy skills in kindergarten with respect to English-language learners and ethnic subgroups. *School Psychology Quarterly, 23*, 553-570. doi:10.1037/1045-3830.23.2.553.

Blagg, H. (2008). *Colonial critique and critical criminology: Issues in Aboriginal Law and Aboriginal violence*. In T. Anthony & C. Cunneen (Eds.), *The critical criminology companion* (pp. 129-145). Sydney: Hawkins Press.

Burgess-Proctor, A. (2006). Intersections of race, class, gender, and crime. Future directions for feminist criminology. *Feminist Criminology, 1*, 27-47. doi:10.1177/1557085105282899.

Canadian Association of Elizabeth Fry Societies. (2008). Submission to the Standing Senate Committee on Legal and Constitutional Affairs Regarding Bill C-2: An Act to amend the Criminal Code and to make consequential amendments to other Acts. Retrieved from http;//www.elizabethfry.ca/risky/Contents.htm

Commonwealth of Australia (2011). *Doing time – time for doing. Indigenous youth in the criminal justice system.* Canberra: Author. Retrieved from http://www.aph.gov.au /house/committee/atsia/sentencing/report/fullreport.pdf

Conway, A. M. (2005). Girls, aggression, and emotion regulation. *American Journal of Orthopsychiatry, 75*, 334-339. doi:10.1037/0002-9432.75.2.334.

Cunneen, C., & White, R. (2007). *Juvenile justice: Youth and crime in Australia* (3rd ed.). South Melbourne, Victoria, Australia: Oxford University Press.

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 and Adolescent Psychiatry, 44,* 798-806. doi:10.1097/01.chi.0000164590.48318.9c.

Dodge, K. A. (2008). Framing public policy and prevention of chronic violence in American youth. *American Psychologist, 63*, 573-590. doi:10.1037/0003-066X.63.7.573.

Eagly, A. H., & Chin, J. L. (2010). Are memberships in race, ethnicity, and gender categories merely surface characteristics? *American Psychologist, 65*, 934-935. doi:10.1037/a0021830.

Embretson, S. E. (1996). The new rules of measurement, *Psychological Assessment, 8,* 341-349.

Emeka, T. Q., & Sorensen, J. R. (2009). Female juvenile risk. Is there a need for gendered assessment instruments? *Youth Violence and Juvenile Justice*, 7, 313-330. doi: 10.1177/1541204009334083.

Farmer, T. W., & Farmer, E. M. Z. (2001). Developmental science, systems of care, and prevention of emotional and behavioural problems in youth. *American Journal of Orthopsychiatry, 71*, 171-181.

Field, A. (2009). *Discovering statistics using SPSS (3rd ed.*). London: Sage Publications.

Fite, P. J., Wynn, P., & Pardini, D. A. (2009). Explaining discrepancies in arrest rates between black and white male juveniles. *Journal of Consulting and Clinical Psychology, 77*(5), 916-927. doi:10.1037/a0016626.

Flores, A. W., Travis, L. F., & Latessa, E. J. (2003). *Case classification for juvenile corrections: An assessment of the YouthLevel of Service/Case Management Inventory (YLS/CMI)*. Cincinnati, OH: Centre for Criminal Justice Research, University of Cincinnati.

Gavazzi, S. M., Yarcheck, C. M., & Chesney-Lind, M. (2006). Global risk indicators and the role of gender in a juvenile detention sample. *Criminal Justice and Behavior, 33(5),* 597-612. doi: 10.1177/0093854806288184.

Gendreau, P., & Smith, P. (2007). Influencing the "people who count": Some perspectives on the reporting of meta-analytic results for prediction and treatment outcomes with offenders. *Criminal Justice and Behavior, 34*, 1536-1559. doi: 10.1177/0093854807307025.

Hanson, R. K., & Howard, P. D. (2010). Individual confidence intervals do not inform decison-makers about the accuracy of risk assessment evaluations. *Law and Human Behavior, 34*, 275-281. doi: 10.1007/s10979-010-9227-3.

Hart, R. K., Michie, C., & Cooke, D. J. (2007). Precision of actuarial risk assessment instruments. Evaluating the 'margins of error' of group v. individual predictons of violence. *British Journal of Psychiatry, 190*, 69-65. doi: 10.1192/bjp.190.5.s60.

Hoge, R. (2008). Assessment in juvenile justice systems. In R. D. Hoge, N. G. Guerra, & P. Boxer (Eds.), *Treating the juvenile offender (pp. 54-75)*. New York: Guilford Press.

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. (2006). *Youth Level of Service/Case Management Inventory. User's Manual*. North Tonawanda, NY: Multi-Health Systems.

Hoge, R. D., & Andrews, D. A. (2011). *Youth Level of Service/Case Management Inventory 2.0. User's Manual*. North Tonawanda, NY: Multi-Health Systems.

Howell, J. C. (1995). *Guide for implementing the comprehensive strategy for serious, violent, and chronic juvenile offenders*. Washington, DC: Office of Juvenile Justice and Deliquency Prevention.

Hua, J., & Fitzgerald, J. (2006). Matching court records to measure reoffending. *Crime and Justice Bulletin: Contemporary Issues in Crime and Justice, no 95.* Sydney, Australia: NSW Bureau of Crime Statistics and Research.

Hubbard, D. J., & Pratt, T. C. (2002). A meta-analysis of the predictors of delinquency among girls. *Journal of Offender Rehabilitation, 34 (3*), 1-13.

Jung, S., & Rawana, E. P. (1999). Risk and need assessment of juvenile offenders. Criminal *Justice and Behavior, 26*, 69-89. doi:10.1177/0093854899026001004

Kehoe, J. K., & Tenopyr, M. L. (1994). Adjustment in assessment scores and their usage: A taxonomy and evaluation of methods. *Psychological Assessment, 6*, 291-303.

Markus, H. R. (2008). Pride, prejudice, and ambivalence: Toward a unified theory of race and ethnicity. *American Psychologist, 63*, 651-670.

Marshall, J., Egan, V., English, M., & Jones, R. M. (2006). The relative validity of psychopathy versus risk/needs-based assessments in the prediction of adolescent offending behaviour. *Legal and Criminological Psychology, 11*, 197-210. doi:10.1348/135532505X68719.

McGrath, A. (in press). The subjective impact of contact with the criminal justice system: The role of gender and stigmatization. *Crime & Delinquency*. doi:10.1177/0011128710389589.

McGrath, A. & Thompson, A. P. (2011). *Actuarial risk assessment for juvenile offenders: Lessons from an Australian normative study*. Paper presented at the 32nd International Congress of Law and Mental Health. Berlin, Germany, 17-23 July.

Medoff, D. (2003). The scientific basis of psychological testing. Considerations following *Daubert*, *Kumho*, and *Joiner*. *Family Court Review, 41(2),* 199-213. doi: 10.1177/1531244503251482.

Mullins, R. L., Cornille, T. A., Mullins, A. K., & Huber, J. (2004). Female juvenile offending: A review of characterisitcs and contexts. *Journal of Child and Family Studies, 13(2),* 205-218. doi: 1062-1024/04/0600-0205/0.

Odgers, C. L., Moretti, M. M., & Reppucci, N. D. (2005). Examining the science and practice of violence risk assessment with female adolescents. *Law and Human Behavior, 29 (1),* 7-27. doi: 10.1007/s10979-005-1397-z.

Odgers, C. L., Robins, S. J., & Russell, M. A. (2010). Morbidity and mortality risk among the "forgotten few": Why are girls in the justice system in such poor health? *Law and Human Behavior, 34*, 429-444. doi:10.1007/s10979-009-9199-3.

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.

Onifade, E., Davidson, W., Campbell, C., Turke, G., Malinowski, J., & Turner, K. (2008). Predicting recidivism in probationers with the Youth Level of Service Case Management Inventory (YLS/CMI). *Criminal Justice and Behavior, 35*,474-483. doi:10.1177/0093854807313427.

Onifade, E., Davidson, W., Livsey, S., Turke, G., Horton, C., Malinowski, J., ... Wimberly, D. (2008). Risk assessment: Identifying patterns of risk in young offenders with the Youth Level of Service/Case Management Inventory. *Journal of Criminal Justice, 36*, 165-173. oi:10.1016/j.jcrimjus.2008.02.006.

Pedersen, A. & Barlow, F. (2008). Theory to social action: A university-based strategy targeting prejudice against Aboriginal Australian. *Australian Psychologist, 43*, 148-159. doi:10.1080/00050060802318587.

Pedersen, A., Walker, I., Paradies, Y., & Guerin, B. (2011). How to cook rice: A review of imngredients for teachng anti-prejudice. *Australian Psychologist, 46*, 55-63. doi:10.1111/j.1742-9544.2010.00015.x.

Phinney, J. S. (1996). When we talk about American ethnic groups, what do we mean? *American Psychologist, 51*, 918-927

Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63*, 737-748.

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.

Rodriguez, N. (2010). The cumulative effect of race and ethnicity in juvenile court outcomes and why preadjudication detention matters. *Journal of Research in Crime and Delinquency, 47*(3), 391-413. doi: 10.1177/0022427810365905.

Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education. Prospects in a post-affirmative-action world. *American Psychologist, 56,* 302-318. doi: 10.1037/0003-066X.56.4.302.

Sackett, P. R., & Wilk, S. L. (1994). Within-group norming and other forms of score adjustment in preemployment testing. *American Psychologist, 49*, 929-954.

Sampson, R. J., Morenoff, J. D., & Raudenbush, S. (2005). Social anatomy of racial and ethnic disparities in violence. *American Journal of Public Health, 95*(2), 224-232.

Schmidt, F., Campbell, M. A., & Houlding, C. (2011). Comparative analyses of the YLS/CMI, SAVRY *Youth Violence and Juvenile Justice, 9*, and PCL:YV in adolescent offenders: A 10-year follow-up into adulthood., 23-42. doi:10.1177/1541204010371793.

Schmidt, F., Hoge, R. D., & Gomes, L. (2005). Reliability and validity analyses of the Youth Level of Service/Case Management Inventory. *Criminal Justice and Behavior, 32*, 329-344. doi:10.1177/0093854804274373.

Schmitt, N., Sackett, P. R., & Ellingson, J. E. (2002). No easy solution to subgroup differences. *American Psychologist, 57*, 305-306.

Schopp, R. F. (1996). Communicating risk assessments. Accuracy, efficacy, and responsibility. *American Psychologist, 51*, 939-944.

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: Predictive validity by gender. *Criminal Justice and Behavior, 35,*1367-1381. doi: 10.1177/0093854808324377.

Schwalbe, C. S., Fraser, M. W., & Day, S. H. (2007). Predictive validity of the joint risk matrix with juvenile offenders: A focus on gender and race/ethnicity. *Criminal Justice and Behavior, 34 (3),* 348-361. doi: 10.1177/0093854806292244.

Schwalbe, C. S., Fraser, M. W., Day, S. H., & Arnold, E. M. (2004). North Carolina Assessment of Risk (NCAR): Reliability and predictive validity with juvenile offenders. *Journal of Offender Rehabilitation, 40*(1/2), 1-22. doi:10.1300/J076v40n01_01.

Schwalbe, C. S., Fraser, M. W., & Day, S. H., & Cooley, V. (2006). Classifying juvenile offenders according to risk of recidivism. Predictive validity, race/ethnicity and gender. *Criminal Justice and Behavior, 33(3),* 305-324. doi: 10.1177/0093854806286451.

Singh, J. P., & Fazel, S. (2010). Forensic risk assessment: A metareview. *Criminal Justice and Behavior, 37*, 965-988. doi:10.1177/0093854810374274.

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. doi:10.1080/00050060500243491.

Thompson, A. P., & Putniņš, A. L. (2003). Risk-need assessment inventories for juvenile offenders in Australia. *Psychiatry, Psychology and Law, 10*, 324-333.

Tolman, A. O., & Mullendore, K. B. (2003). Risk evaluations for the courts: Is service quality a function of specialization? *Professional Psychology: Research and Practice, 34 (3),* 225-232. doi: 10.1037//0735-7028.34.3.225.

Tolman, A. O., & Rotzien, A. L. (2007). Conducting risk evaluations for future violence: Ethical practice is possible. *Professional Psychology: Research and Practice, 38,* 71-79. doi:10.1037 /0735-7028.38.1.71.

Tonry, M., & Melewski, M. (2008). The malign effects of drug and crime control policies on black Americans. In M. Tonry (Ed.), *Crime & Justice: A Review of Research* (Vol. 37, pp. 1-44). Chicago: University of Chicago Press.

Weatherburn, D., Snowball, L., & Hunter, B. (2008). Predictors of Indigenous arrest: An exploratory study. *Australian & New Zealand Journal of Criminology, 41*, 307-322. Doi: 10.1375/acri.41.2.307.

Table 1
Mean (standard deviation) for the YLS/CMI-AA scores by gender and by ethnic subgroups

	Gender		Australian Ethnicity		
YLS/CMI-AA domain	Female (n= 559)	Male (n= 3009)	Indigenous (n = 1053)	Australian (n= 1582)	Ethnic (n = 765)
Prior/current offences	3.50 (1.67)	3.58 (1.78)	4.18_a (1.82)	3.51_b (1.73)	3.10_c (1.56)**
Family and living	2.83 (1.90)	2.26 (1.89)*	2.77_a (1.90)	2.41_b (1.92)	1.80_c (1.78)**
Education/employment	2.57 (2.10)	2.58 (2.16)	2.88_a (2.19)	2.64_b (2.18)	2.24_c (1.99)**
Peer relations	2.17 (1.37)	2.13 (1.34)	2.59_a (1.28)	2.00 (1.36)	1.92_b (1.24)**
Substance abuse	2.55 (1.91)	2.27 (1.99)*	2.64_a (2.06)	2.51 (1.95)	1.58_b (1.76)**
Leisure/recreation	1.70 (1.14)	1.49 (1.18)*	1.75_a (1.17)	1.52_b (1.16)	1.32_c (1.20)**
Personality/behavior	2.38 (2.14)	2.08 (2.07)*	2.38_a (2.15)	2.23 (2.13)	1.69_b (1.86)**
Attitudes/beliefs	1.19 (1.48)	1.13 (1.44)	1.38_a (1.54)	1.14_b (1.46)	0.92_c (1.31)**
Total score	18.90 (9.52)	17.54 (9.66)*	20.57_a (9.54)	17.95_b (9.67)	14.57_c (8.69)**

Note. For ethnic subgroups, means with different subscripts differ significantly. *$p < .005$. **$p < .001$.

Table 2

One Year Predictive Validity (Recidivism) for YLS/CMI-AA Total Score by Subgroups

			ROC Analysis		
Group	*n*	*r*	*AUC*	*SE*	95% CI [LL, UL]
Female	559	.24**	.644***	.023	[.598, .690]
Male	3009	.27**	.659***	.010	[.639, .678]
Indigenous	1053	.17**	.604***	018	[.569, .639]
Australian	1582	.25**	.644***	.014	[.617, .671]
Ethnic	765	.26**	.652***	.020	[.613, .692]
Total	3568	.26**	.652***	.009	[.634, .670]

** $p < .01$;*** $p < .001$

Table 3

One year reoffending number (percent) for YLS/CMI-AA risk categories by subgroups

	Gender			Australian Ethnicity			
Risk category	Female (n= 559)	Male (n= 3009)	Combined[a] (n = 3568)	Indigenous (n = 1053)	Australian (n = 1582)	Ethnic (n = 765)	Combined[a] (n = 3400)
Low	64 (29.4)	561(39.8)	625 (38.4)	186 (53.0)	270 (38.1)	142 (31.3%)	598 (39.5)
Moderate	123 (44.9)	780(61.2)	903 (58.3)	359 (66.6)	383 (55.6)	141 (52.0)	883(58.9)
High	39 (58.2)	242(74.2)	281 (71.5)	123 (75.5)	128 (69.6)	28 (68.3)	279 (71.9)
All[b]	226 (40.4)	1583 (52.6)	1809 (50.7)	668 (63.4)	781 (49.4)	311 (40.7)	1760 (51.8)

[a] Sample size for combined groups for gender and ethnicity differ due to the exclusion of cases with missing or unknown ethnicity data,

[b] Offending rate for all risk categories is base rate per subgroup

Note. YLS/CMI-AA total score for risk categories is Low risk (0 to 15), Moderate risk (16 to 30), High Risk (31-48)