Proceedings
of the 2006 Joint Conference of the APS
and NZPsS
26 – 30 September 2006, Auckland, NZ

Psychology Bridging the Tasman: Science,
Culture and Practice

Editor: Dr Mary Katsikitis

ISBN 0-909881-30-8

Publisher
The Australian Psychological Society Ltd
Level 11, 257 Collins Street, Melbourne, Victoria 3000 Australia
Email: proceedings@psychology.org.au
Do We Need to Worry About Subgroup Differences in Juvenile Offender Risk-need Assessment?

Anthony P. Thompson (athomspn@csu.edu.au)
School of Humanities & Social Sciences
Charles Sturt University, Wagga Wagga NSW 2678 Australia

Abstract

Since 2002 the Department of Juvenile Justice (NSW) has been using the Australian Adaptation of the Youth Level of Service/Case Management Inventory (Hoge & Andrews, 1995) as part of the routine assessment of young offenders. Over three years, 6802 inventories were completed by juvenile justice officers. Although risk-need assessment is common in the juvenile and adult criminal justice systems, some observers worry about test bias for particular subgroups. The NSW database was used to examine subgroup differences for risks, needs and strengths between females (N=644) and males (N=3594) and between Aboriginal (N = 1320) and Australian-Non-Aboriginal (N=1848) young people. Significant differences are interpreted in terms of effect size, item analysis, and issues that guide interpretation and use of risk-need scores.

Introduction

Recognising human diversity is a strong element of modern day psychology with its emphasis on multi-cultural competence (Stuart, 2004). Human diversity extends beyond culture to include any group differences that are salient factors for assessment and intervention (Roysircar, 2004). The pursuit of subgroup differences though requires sensitivity. Stereotype and prejudice are readily aroused (Pliner & Sakamoto, 2005) as they have long been in the troubled course of research into subgroup differences in intelligence (Gordon, 1995). The current study investigated subgroup differences in risk-need assessment with juvenile offenders. The emphasis was not just on determining whether differences exist, but on understanding the basis of any difference and implications for professional practice.

Assessing the risk factors and intervention needs of juvenile offenders is widely recognised as a best practice approach (Day, Howells, & Rickwood, 2004; McLaren, 2000, Thompson, 2001). Such assessment is part of a broader criminal justice strategy which involves managing the risk of repeat offending and at the same time responding to the psychosocial needs of adolescents. To that end, risk-need assessment inventories have been developed in Canada (Hoge & Andrews, 2002), the United States (Howell, 1995; Schwalbe et al. 2004), England and Wales (Baker, Jones, Roberts, & Merrington, 2003) and Australia (Puttnies, 2005; Thompson & Pope, 2005).

Risk-need inventories are essentially structured checklists based on the major psychosocial risk factors for juvenile offending that have emerged from an extensive empirical and theoretical literature (Thompson & Puttnies, 2003). Typically, inventories also incorporate variables related to prior history of offending which have proved useful in predicting recidivism. There are an increasing number of publications on risk-need inventories for juvenile offenders. Most often, these report on development of the inventory and its psychometric properties. There has also been research into predictive validity. Generally, inventory scores are found to bear a modest relationship with re-offending (e.g., Puttnies, 2005; Schmidt, Hoge, & Gomes, 2005; Thompson & Pope, 2005).

Less attention has been directed at whether identifiable subgroups score differently on such inventories, although some results are available. For example, Jung and Rawana (1999) compared Native with Non-Native juvenile offenders and males with females in a Canadian sample. The Native sub-sample scored significantly higher overall on the risk-need scale but this only amounted to a difference of approximately 2.5 points on a 43 point scale. This was attributable to more difficulties with negative peers, substance use and leisure time. There were no differences between males and females. Regardless of ethnicity or gender, the risk-need score discriminated between recidivists and non-recidivists. Using the same inventory, but in a smaller and more restrictive sample, Schmidt et al. (2005) investigated gender subgroups and found essentially the same results as Jung and Rawana.

Schwalbe, Fraser, Day, & Arnold, (2004) emphasise that the predictive validity of risk assessment instruments may vary for diverse populations. This conclusion was based on studying the relationship between risk instrument scores and recidivism over a 12 month period. They found different non-linear relationships for gender and ethnic subgroups (European White American vs. African American). They also found differences in the total risk score for subgroups defined by substance use (non-users, non-abusive users and abusers). Other research (National Council on Crime and Delinquency, 2000; Puttnies, 2005) also supports the view that ethnic and gender
subgroups may score differently and validity coefficients may similarly vary.

In summary, risk-need inventories are used widely in juvenile justice settings and their application must be informed by ongoing research and evaluation. An important question concerns whether specific subgroups differ on such instruments. This question is important for several reasons. First, it is accepted that the precursors of crime and the dynamics of the criminal justice system are different depending on characteristics such as gender and racial/ethnic status (Cummins & White 2002; Homel, Lincoln, & Herd, 1999). Secondly, practitioners can benefit from being sensitised to risk-need considerations that may be particularly relevant for certain groups. Thirdly, the spectre of test bias must be considered for both psychometric and human rights reasons (Canadian Association of Elizabeth Fry Societies, n.d.; The Australian Psychological Society Limited, 2003).

**Method**

This study used a data set provided by the Department of Juvenile Justice, NSW. In 2002, the Department adopted a risk-need inventory as a routine element of its assessment procedures with all juvenile offenders. The inventory, its use and the data set are described next.

**Australian Adaptation of the Youth Level of Service/ Case Management Inventory (YLS/CMI-AA, Hoge & Andrews 1995)**

This inventory provides a structured approach to assessing the risk, needs and strengths of juvenile offenders. The Australian adaptation began prior to the commercial availability of the parent inventory (Hoge & Andrews, 2002). Details on the adaptation and its preliminary psychometric properties are provided by Thompson & Pope (2005). As with the parent version, items in the Australian trial version are organised into the following domains: (1) Prior and Current Offences (eight items), (2) Family and Living Circumstances (seven items), (3) Education/Employment (seven items), (4) Peer Relations (four items), (5) Substance Abuse (six items), (6) Leisure/Recreation (three items), (7) Personality/Behaviour (seven items), (8) Attitudes/Orientation (five items). A new domain, Assessment of Major Strengths, was included for three items related to protective factors.

Items are scored in a binary fashion to indicate whether the operationally defined item applies to the young person. One item related to age at first court order was scored 0, 1, and 2 with more weight given to younger offenders. Major strengths are indicated as present or not at the individual level (social and personal skills), family level (strong, positive parent-child relationship) and community level (support outside the family). Items are tallied to provide domain totals. An overall score (range 0 – 48) is calculated based on all domains except “Major Strengths”.

**Procedure for Completing the Inventory**

The YLS/CMI-AA is intended for use by professionals working in the juvenile justice sector. In NSW it is used in computerised format under a licensing agreement with the test publisher. Juvenile Justice Officers (JJOs) complete the inventory for all young persons involved in supervision or control orders. Use of the inventory is supported by staff training and a departmental procedure manual. The inventory is part of the initial and ongoing assessment of clients and is linked to case management. Guidelines require the inventory to be completed based on interview and collateral information. JJOs enter item responses via their computer which produces scores and a profile. The results are used to support decisions about level of intervention and case planning. The inventory is completed again if there are significant changes in the young person’s situation or after 6 months if the previous ratings no longer apply.

**The Data Set**

Computerised use of the inventory allows for immediate inclusion of results into the Department’s Client Information Management System. This is advantageous for ongoing evaluation and related research. Electronic data for YLS/CMI-AA inventories completed between November 2002 and November 2005 were provided to the author. Ethics approval for the research was granted through ethics committees at the university and Department of Juvenile Justice. In raw form, there were 6890 data sets from 41 juvenile justice locations throughout the state.

**Results**

Although there were 6890 sets of data, a smaller number of individuals are represented as repeat assessments are included. The database was restructured by client identification number. Visual and electronic inspection resulted in the removal of 1.28% of entries which were duplicated or incomplete. The remaining 6802 inventory results arose from 4297 separate juvenile offenders. Of these 2628 (61%) had only 1 inventory completed, 1079 (25%) had two inventories completed and 590 (14%) had three or more inventories completed. Further refinement of the data set was undertaken to remove 59 cases (1.4%) in which the second set of inventory results were completed within 30 days of the first. For many cases, it appeared that the two sets arose from data submission errors such as submitting results twice or submitting an incomplete inventory. Thus, the results reported below arise from inventories completed for 4238 juveniles.
Table 1. Mean (standard deviation) for the YLS/CMI-AA scores by gender and ethnic subgroups

<table>
<thead>
<tr>
<th>YLS/CMI-AA</th>
<th>Gender</th>
<th>Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female (N = 630)</td>
<td>Australian (1299)</td>
</tr>
<tr>
<td></td>
<td>Male (N=3472)</td>
<td>Non-ATSI (1792)</td>
</tr>
<tr>
<td>Prior and current</td>
<td>3.65 (1.77)</td>
<td>4.46 (1.94)</td>
</tr>
<tr>
<td></td>
<td>3.79 (1.94)</td>
<td>3.65 (1.86)</td>
</tr>
<tr>
<td>Family and living</td>
<td>2.93 (1.94)</td>
<td>2.94 (1.93)</td>
</tr>
<tr>
<td></td>
<td>2.43 (1.94)</td>
<td>2.52 (1.96)</td>
</tr>
<tr>
<td>Education</td>
<td>2.62 (2.09)</td>
<td>2.95 (2.19)</td>
</tr>
<tr>
<td></td>
<td>2.72 (2.19)</td>
<td>2.76 (2.20)</td>
</tr>
<tr>
<td>Peer relations</td>
<td>2.20 (1.39)</td>
<td>2.68 (1.28)</td>
</tr>
<tr>
<td></td>
<td>2.24 (1.35)</td>
<td>2.09 (1.39)</td>
</tr>
<tr>
<td>Substance abuse</td>
<td>2.64 (1.93)</td>
<td>2.83 (2.10)</td>
</tr>
<tr>
<td></td>
<td>2.44 (2.04)</td>
<td>2.63 (1.99)</td>
</tr>
<tr>
<td>Leisure/recreation</td>
<td>1.72 (1.13)</td>
<td>1.82 (1.16)</td>
</tr>
<tr>
<td></td>
<td>1.57 (1.19)</td>
<td>1.59 (1.17)</td>
</tr>
<tr>
<td>Personality</td>
<td>2.48 (2.15)</td>
<td>2.55 (2.17)</td>
</tr>
<tr>
<td></td>
<td>2.24 (2.11)</td>
<td>2.34 (2.16)</td>
</tr>
<tr>
<td>Attitudes</td>
<td>1.28 (1.51)</td>
<td>1.51 (1.57)</td>
</tr>
<tr>
<td></td>
<td>1.25 (1.50)</td>
<td>1.24 (1.50)</td>
</tr>
<tr>
<td>Total score</td>
<td>19.51 (9.74)</td>
<td>21.72 (9.70)</td>
</tr>
<tr>
<td></td>
<td>18.68 (10.07)</td>
<td>18.82 (10.10)</td>
</tr>
<tr>
<td>Strengths</td>
<td>1.59 (0.99)</td>
<td>1.56 (1.05)</td>
</tr>
<tr>
<td></td>
<td>1.64 (1.04)</td>
<td>1.58 (1.02)</td>
</tr>
</tbody>
</table>

Gender Comparisons

In the sample there were 644 (15.2%) females and 3594 (84.8%) males. The mean age for females was 16.35 years compared with 16.56 years for males. Descriptive statistics for the YLS/CMI-AA domain scores and the overall total score are summarised in Table 1.

A multivariate analysis of variance (MANOVA) was performed to investigate gender differences in domain risk-need scores (excluding the strengths domain). Violation of some statistical assumptions, such as normality of variables and equality of error variances, was mitigated by large sample size and use of conservative alpha levels (Pallant, 2005). There was a significant difference between females and males on the combined dependent variables: $F(8, 4093) = 11.61$, $p < .001$; Wilks’ Lambda = .98; partial eta squared = .02. Considering dependent variables separately, the only difference to reach statistical significance, using an alpha level of .001, was “Family and Living Circumstances”: $F(1,4100) = 36.06$, $p < .001$; partial eta squared = .01. This is considered a small effect. Females were rated to have slightly more family problems ($M = 2.93$) than males ($M = 2.45$). Analysis of the 7 items in this domain showed endorsement proportions significantly higher ($p < .01$) for females on all items except one concerning antisocial values in the family. Endorsement proportions varied 4%-7% except for the item related to poor relations with mother or step-mother which had a 13.7% gender differential.

An independent sample t-tests on the total risk-need score was not significant ($t(4100) = 1.92, p = .06$). To compare the distribution of total scores, each score was categorised as low (0-11), moderate (11-23), high (24-36) and very high (37-48), similar to the categorizations used by Hoge and Andrews (2002). The categorical distributions were not significantly different ($p < .01$) for females and males. Gender difference on the mean score for “Major Strengths” was not significant ($t(4236) = 1.04, p = .30$). However, at the item level, more females (62.3%) than males (56.1%) were judged to have individual strengths ($Pearson \chi^2(1) = 8.10, p < .005$) but fewer females (55.9%) than males (66.4%) were judged to have “family” strengths ($Pearson \chi^2(1) = 26.17, p < .001$).

Ethnic Comparisons

In the sample there were 1320 (31%) young persons who self-identified as being of Aboriginal or Torres Strait Islander (ATSI) ethnic origin. There were 1848 (44%) who self-identified as being of Australian (Non-ATSI) background. Another 920 (22%) identified one of 91 other ethnic backgrounds. The main focus of the analysis was the ATSI and Non-ATSI sub-samples. The mean age of the ATSI offenders was 16.18 years compared with 16.61 years for the Non-ATSI offenders. Descriptive statistics are in Table 1.

A MANOVA investigated ethnic differences in domain risk-need scores. Again, some statistical assumptions were violated but mitigated by large sample size and conservative alpha levels. There was a significant difference between ATSI and Non-ATSI samples on the combined dependent variables: $F(8, 3082) = 28.94, p < .001$; Wilks’ Lambda = .93; partial eta squared = .07. Most of the dependent variables were significantly different ($p < .01$) between the two groups, but for only two domains was the effect size moderate. Mean scores indicated that the ATSI sample scored higher on “Prior and Current Offences” ($F(1, 3089) = 134.28, p < .001$; partial eta squared = .042) and on “Peer Relations” ($F(1, 3089) = 142.58, p < .001$; partial eta squared = .044). Item analyses in these two domains
showed significant group differences (p < .001) in endorsement proportions for most items. Largest differences showed Aboriginal youth more likely to: 1) receive a court order when they were 14 years or younger (40% vs. 25%), 2) have three or more prior offences (41% vs. 27%) and 3) have a conviction for common assault, break and enter or motor vehicle theft (74% vs. 61%). Aboriginal youth were also judged more often to have negative peer influences with all four items in this domain having a 13–18% differential.

On the overall risk-need score, the ATSI group scored higher than the Non-ATSI group (t (3089) = 8.01, p < .001) which is a small effect size. Analysis of total score categorizations (defined previously) showed a significant difference between ATSI and Non-ATSI groups (Pearson χ² (3) = 67.96, p < .001). Fewer in the ATSI group were in the low category (14% vs. 25%), but more were in the high category (33% vs. 26%) and very high category (8.7% vs. 5.7%). Between group mean difference on “Major Strengths” was not significant (t (3166) = 0.49, p = .63). At the item level though, ATSI were less frequently judged to have individual strengths (49% vs. 58%) but more often judged to have community strengths (45% vs. 37%).

Other Analyses
It was important to investigate the possibility of interaction effects when gender and ethnicity are considered jointly. An additional MANOVA of domain scores with gender (male, female) and ethnicity (ATSI, Non-ATSI) as fixed factors revealed no overall interaction effect (F (8,3080) = 1.06, p = .39; Wilks’ Lambda = 1.00). Similarly, there was no significant interaction for the inventory total score or the score for the strength domain. Thus, the main findings related to gender hold for ethnic sub-groups and vice versa.

Previous analyses focused on ATSI and Non-ATSI subgroups, but there were 920 offenders from other ethnically diverse backgrounds. As a group, these offenders were distinctly different from the ATSI and Non-ATSI groups. For example, they were older (16.80 years), had a lower overall YLS/CMI-AA score (15.40), lower scores on all risk-need domains and a higher mean score on the Strength domain. All differences were significant (p < .001) using appropriate multivariate/univariate analyses and post hoc tests.

Discussion
The results show little reason to be concerned about gender differences on the risk-need inventory. The overall score and distribution of categorical ratings were not significantly different. The only difference worth noting is that females were judged to have more problems in the family domain. However, this effect was quite small amounting to an overall difference of half a point for the domain total. Conceptually though, this results is consistent with literature on female offending. Mullis, Cornille, Mullis, & Huber, (2004) identified family relations as a key risk factor for female delinquency. Family problems may arise from the high incidence of abuse among female offenders, poor parenting practices and emotional conflict. Resulting family alienation can lead to anti-social peers and accommodation problems. Although some suggest that there are distinctive family problems for female offenders, a meta-analysis by Simour and Andrews (1994) concluded that poor-parent child relations was an important risk factor for both genders. At a broader level, Cunneen and White (2002) stress the influence of systemic factors and refer to the criminalization of welfare needs in the treatment of young women. Based on the results of the current data, it would be prudent for practitioners using the risk-need inventory to: 1) be aware that gender based score differentials are minimal 2) explore in detail issues related to family and living circumstances that may be particularly salient for young female offenders.

The analyses of this study showed more compelling differences between ATSI and Non-ATSI offenders on the risk-need inventory. Nevertheless, statistically the effect is regarded as small. Overall, the ATSI group mean was 3 points higher. Many would not be surprised that this difference was largely attributable to higher scores due to past and current convictions and to peer relations. More debatable is the explanation and implication of this difference. One view is that the criminal justice system is applied more forcefully to Aboriginals leading to earlier contact for youth, less use of diversionary procedures and more serious sanctions. With more criminal convictions, Aboriginal youth socializing with each other will have more peers with a criminal history. An alternative view is that certain forms of criminality are more prevalent among Aboriginal youth due to colonial history, social and economic problems and cultural traditions.

The fact that scores are different on the YLS/CMI-AA for ATSI and Non-ATSI subgroups is not in itself evidence of test bias. Test bias is a complex issue that technically hinges on the relationship between test scores and outcome measures (Sackett & Wilk, 1994). If the predictive relationship is different, then use of test scores can result in disadvantage although sometimes in counter intuitive ways. Currently, the relationship between the YLS/CMI-AA and recidivism is being investigated for gender and ethnic sub-groups. In the mean time, practitioners should understand rather than worry about subgroup differences. In many jurisdictions, risk-need scores are used predominantly for case planning rather than for what is referred to as “high stakes” decisions (Sackett et al., 2001). Knowledge of subgroup differences can help
contextualize inventory results. In addition, using score bands for the YLS/CMI-AA total rather than using raw scores can help minimize the impact of subgroup differences and over interpretation of small score differences. It is also worth stressing that risk-need inventories permit the kind of transparent evaluation represented by the results reported here. Less structured evaluation of juvenile offenders may be prone to more dramatic subgroup differences.

Acknowledgments

The author thanks the NSW Department of Juvenile Justice for assistance in undertaking the research. The opinions here do not necessarily reflect the views of the NSW Department of Juvenile Justice, or any of its officers.

References


Contents.htm


