The internal reliability and construct validity of the Offender Assessment System (OASys)
Robin Moore

The aim of the research was to assess the internal reliability and construct validity of the core component of the Offender Assessment System (OASys), the national risk and need assessment tool for adult offenders in England and Wales. In its current format, the core component has 73 scored questions across eleven scales, covering offending information and a range of individual-level and social factors. The total OASys score has reasonable predictive validity, and two new predictors have recently been developed for general re-offending and violent re-offending. This paper assesses two further aspects of reliability and validity: (a) **internal reliability** – how well the items within each scale measure various aspects of the same characteristic; and (b) **construct validity** – how well the assessment tool distinguishes between discrete individual-level or social characteristics. Attention is then given to improving the measurement of discrete criminogenic needs. While the ‘What Works’ risk principle requires correspondence between the intensity of interventions and offenders’ risk of re-offending levels, the criminogenic need principle requires that, on the grounds of efficiency and effectiveness, interventions should be targeted towards dynamic and changeable criminogenic needs.

**Key points**

- Analysing approximately 230,000 valid OASys assessments completed during 2006/07, six of the eleven scored OASys scales were found to have high internal reliability, clearly measuring a discrete characteristic, and four had adequate internal reliability. ‘Relationships’ was the only scale with non-adequate internal reliability, indicating that the questions within the section were failing to measure a single factor.

- The construction of OASys could be improved through a reduction from 73 scored questions across eleven scales to 47 scored questions across the ten individual-level and social problem scales (sections 3 to 12).

- While the revisions would ensure that the ten scales were measuring distinct problem areas, three of the scales – the relationships scale, the lifestyle and associates scale and the emotional wellbeing scale – were not found to be measuring independently significant ‘criminogenic’ needs.

- The cut-off points for identifying criminogenic needs should be set in relation to re-offending rates. Adjustments in the allocation of resources would be required to ensure that interventions were available to address the revised criminogenic need levels. Offenders with ‘high’ levels of need should be distinguished from offenders with ‘medium’ levels of need to assist with the targeting of interventions, maximising the use of resources.

- The optimum criminogenic need cut-off points for different age and gender subgroups should be recalculated once larger samples are available. Any widening in the targeting of OASys would increase the validity of the calculations to the complete prison and probation caseloads.
• Across sections 3 to 12 of OASys, nine of the currently scored questions are not needed within the revised individual-level or social problem scales or the new violent and general re-offending predictors. These questions could be removed from OASys unless: (a) they are found to be helpful in assessing risk of serious harm; (b) they serve another specific purpose for practitioners; and/or (c) further research reveals that they could be improved through amendments to their wording or accompanying guidance.

• The potential value of additional questions should be considered following an evaluation of the textual information recorded by assessors within each of the OASys sections. The initial focus should be placed upon the relationships, the lifestyle and associates and the emotional wellbeing sections, identifying alternative questions which are amenable to change and have stronger independent associations with re-offending.

Approach
The main sample consisted of 230,163 OASys assessments administered by probation assessors during 2006/07. All 73 scored questions had to have been completed, and the sample was restricted to one assessment per offender. The sample should not be seen as representative of the entire offender population, with previous analysis having found that offenders with an assessment are more likely to have committed a violent offence and to have a high likelihood of reconviction. To guide the analysis, the following three research questions were set.

• Do the scored questions within each scale measure discrete individual-level or social characteristics? (Internal reliability)

• What are the common factors underlying the scored questions? (Construct validity)

• What improvements could be made to the assessment of criminogenic needs? (Improving OASys)

To evaluate internal reliability, Cronbach’s alpha scores were used, measuring how well the individual questions in each OASys scale correlated with the sum of the remaining questions. Item-scale correlations were also calculated. To evaluate construct validity, factor analysis was used, assessing the interrelationships among all scored questions and then explaining these questions in terms of their common underlying dimensions (factors). To look at the relationships with re-offending and to identify the most appropriate criminogenic need cut-off points, a further sample of 43,695 valid OASys assessments from late 2004/early 2005 with 24-month re-offending data was analysed. Logistic regression was used to account for the correlations between the scales, with odds ratios being used to establish cut-off points, comparing the odds of re-offending for offenders with a particular score to the average odds of re-offending.

Results
Using Cronbach’s alpha cut-off points of 0.7 for adequate scores and 0.8 for high scores, six of the eleven OASys scales had high reliability, clearly measuring discrete characteristics, and four scales had adequate reliability. ‘Relationships’ was the only scale with non-adequate reliability. ‘Relationships’ was the only scale with non-adequate reliability. In other words, the questions within this section were failing to measure a single factor. For six of the 73 questions, the scale’s alpha score increased when it was removed, suggesting that it was not contributing to the scale’s internal reliability. Four questions had low item-scale correlations (less than 0.3), three of which were from the relationships section.

Factor analysis revealed 15 underlying factors explaining 60% of the variation in the variables. Eight of the factors corresponded to the current OASys scales, with the other three scored scales each being divided into two factors. The final factor focused upon violence, comprising of three questions from two different OASys scales. Concentrating upon the underlying criminogenic needs covered by sections 3 to 12 of OASys, and removing those questions which: (a) had a detrimental impact in terms of internal reliability and/or low item-scale correlations; and/or (b) fell within non-corresponding factors, 47 questions were then grouped into ten factors in accordance with their current scales. As previously, the relationships scale had non-adequate reliability, although the Cronbach’s alpha had risen to above 0.6, while the alpha score for lifestyle and associates fell to below 0.7. Alpha scores are a function of the number of scale items as well as the item correlations, and both
these scales had just three remaining questions. When three of the 47 questions were removed, the relevant scale’s alpha score increased, but these increases were minimal. None of the remaining questions had an item-scale correlation below 0.3.

Using logistic regression to account for the relationships between the scales and the Offender Group Reconviction Scale (OGRS) 3 score (based upon static criminal history and offender demographic factors), the following three scales were not found to be independently associated with re-offending: (a) the relationships scale; (b) the lifestyle and associates scale; and (c) the emotional wellbeing scale. In other words, while these three scales were measuring distinct problem areas, they were not measuring independent criminogenic needs. Optimum criminogenic need cut-off points for the seven remaining scales were calculated on the basis of odds ratios, comparing the odds of re-offending for offenders with a particular score to the average odds of re-offending. Across all seven scales, there was a point at which the odds ratio increased to a value greater than one, i.e. where the odds of re-offending exceeded the average odds of re-offending. Offenders scoring above this point were judged to have ‘medium’ levels of need. There was also a point across all of the scales, except for the accommodation scale, at which the odds ratio exceeded two, enabling a distinction to be made between offenders with ‘high’ levels of need and those with ‘medium’ levels of need. Table 1 sets out the revised score ranges and criminogenic need cut-off points.

Table 2 demonstrates that the odds of re-offending for those with high levels of need were at least twice the odds of re-offending for those with none/low levels of need, with an odds ratio greater than four for the drug misuse scale and an odds ratio greater than three for four of the other scales. Distinguishing between offenders with none/low levels of need and those with high levels of need produced greater odds ratios than those produced by both the current criminogenic need yes/no scales and the assessors’ own judgements regarding links to offending behaviour. The correlations with re-offending, measured through phi coefficients, were higher than those produced by the assessors’ own judgements across all scales and at least as high as those produced by the current criminogenic need scales across six of the revised scales.

Comparing the criminogenic need prevalence rates produced by the revised scales (including offenders with medium or high levels of need) to those produced by the current scales, there were increases in prevalence rates across four of the seven scales. By far the greatest change was a 31.8% increase for financial management and income; the greatest negative change was a 15.3% fall in alcohol misuse. Adjustments in the allocation of resources would thus be required to ensure that interventions were targeted at criminogenic needs. Comparing those offenders with high levels of need under the revised system to those with identified needs under the current system, the prevalence rates decreased across all scales. It would, therefore, be possible to target interventions at more discrete groups of offenders than currently identified.

<table>
<thead>
<tr>
<th>OASys scale</th>
<th>Cronbach’s alpha</th>
<th>No. scored questions</th>
<th>Score range</th>
<th>None/low need / medium need cut-off</th>
<th>Medium need / high need cut-off</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accommodation</td>
<td>.937</td>
<td>4</td>
<td>0 – 8</td>
<td>1+</td>
<td>-</td>
</tr>
<tr>
<td>Education, training and employability</td>
<td>.817</td>
<td>6</td>
<td>0 – 12</td>
<td>4+</td>
<td>9+</td>
</tr>
<tr>
<td>Financial management</td>
<td>.796</td>
<td>4</td>
<td>0 – 8</td>
<td>2+</td>
<td>7+</td>
</tr>
<tr>
<td>Drug misuse</td>
<td>.808</td>
<td>5</td>
<td>0 – 10</td>
<td>1+</td>
<td>5+</td>
</tr>
<tr>
<td>Alcohol misuse</td>
<td>.881</td>
<td>5</td>
<td>0 – 10</td>
<td>6+</td>
<td>10</td>
</tr>
<tr>
<td>Thinking and behaviour</td>
<td>.843</td>
<td>6</td>
<td>0 – 12</td>
<td>5+</td>
<td>11+</td>
</tr>
<tr>
<td>Attitudes</td>
<td>.716</td>
<td>5</td>
<td>0 – 10</td>
<td>2+</td>
<td>4+</td>
</tr>
</tbody>
</table>
Table 2: 24-month re-offending rates by revised criminogenic need cut-offs

<table>
<thead>
<tr>
<th>OASys scale</th>
<th>% re-offended</th>
<th>Phi coefficient</th>
<th>Odds ratio (significant need vs. no need)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>None/low need</td>
<td>Medium need</td>
<td>High need</td>
</tr>
<tr>
<td>Accommodation</td>
<td>38.4%</td>
<td>55.4%</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>171</td>
<td>.261</td>
<td>3.80</td>
</tr>
<tr>
<td>ETE</td>
<td>34.1%</td>
<td>55.3%</td>
<td>66.3%</td>
</tr>
<tr>
<td></td>
<td>.261</td>
<td>3.80</td>
<td></td>
</tr>
<tr>
<td>Financial management</td>
<td>37.1%</td>
<td>53.4%</td>
<td>67.5%</td>
</tr>
<tr>
<td></td>
<td>.186</td>
<td>3.53</td>
<td></td>
</tr>
<tr>
<td>Drug misuse</td>
<td>36.1%</td>
<td>57.1%</td>
<td>71.3%</td>
</tr>
<tr>
<td></td>
<td>.276</td>
<td>4.40</td>
<td></td>
</tr>
<tr>
<td>Alcohol misuse</td>
<td>43.4%</td>
<td>53.2%</td>
<td>64.1%</td>
</tr>
<tr>
<td></td>
<td>.097</td>
<td>2.35</td>
<td></td>
</tr>
<tr>
<td>Thinking &amp; behaviour</td>
<td>35.8%</td>
<td>54.7%</td>
<td>65.7%</td>
</tr>
<tr>
<td></td>
<td>.204</td>
<td>3.43</td>
<td></td>
</tr>
<tr>
<td>Attitudes</td>
<td>35.0%</td>
<td>53.3%</td>
<td>68.0%</td>
</tr>
<tr>
<td></td>
<td>.262</td>
<td>3.95</td>
<td></td>
</tr>
</tbody>
</table>

Implications

While the core OASys assessment has reasonable construct validity, not all questions load onto the correct factors, and the analysis of each scale’s internal reliability revealed that not all questions are contributing to the measurement of discrete individual-level or social problems. An opportunity is thus available to streamline the assessment, while at the same time maintaining its predictive validity, helping to identify which offenders should receive the available interventions (the risk principle), and improving its internal reliability and construct validity, helping to identify which problems should be addressed (the criminogenic need principle). Recognising that there is no longer any requirement to produce a score for offending information (following the development of the new OASys re-offending predictors), the analysis supports a reduction from 73 scored questions across eleven scales to 47 scored questions across the ten individual-level or social problem scales. Seven of these scales were found to be measuring not only distinct problem areas but independently significant criminogenic needs.

Across sections 3 to 12 of OASys, those nine questions that are not used either within the revised scales or the new re-offending predictors should only be retained if: (a) they are found to be helpful in assessing risk of serious harm; (b) they serve another specific purpose for practitioners; and/or (c) further research, e.g. the probation inter-rater reliability study, reveals that they could be improved through amendments to their wording or accompanying guidance. Those seven questions used within the new predictors but not the revised scales should remain in their current sections without contributing to the identification and measurement of criminogenic need.

The criminogenic need cut-off points should be set on the basis of odds ratios, comparing the odds of re-offending for offenders with a particular score to the average odds of re-offending. Offenders with ‘high’ levels of need should be distinguished from offenders with ‘medium’ levels of need, assisting with the targeting of interventions and maximising the use of resources. The optimum criminogenic need cut-off points for different age and gender subgroups should be recalculated once larger samples are available. Any widening in the targeting of OASys would increase the validity of the calculations to the complete prison and probation caseloads.

The potential value of additional questions should be considered following the completion of research looking at the textual information recorded by assessors within each of the OASys sections, identifying alternative questions which are amenable to change and have stronger independent associations with re-offending. Emphasis should be placed upon the relationships, the lifestyle and associates and the emotional wellbeing sections. The revised scales for these three sections were not found to be measuring distinct criminogenic needs, despite evidence within the ‘What Works’ literature that the first two represent central risk factors.

Note: A full report of this research is included in the OASys Data Evaluation and Analysis Team - Compendium of research and analysis on the Offender Assessment system 2006-2008.