Predictive Validity of the YLS/CMI In Nebraska

Dr. Richard L. Wiener
Law/Psychology Program - UNL

Funding Source: Title II award from the Nebraska Crime Commission

rwiener2@unl.edu
The **Risk-Need-Responsivity Model (RNR)** (Andrews and Bonita, 2010)

- Assess risk through criminogenic needs

- Intervene through techniques that are Evidence Based and that are tailored to the characteristics of the offender
RNR

Assessing Risk
Evidence Based Practice
Tailoring to Individual Needs
YLS/CMI
(Hoge and Andrews, 2002)

• Adaptation of the LSI-R
• Measures risk and needs of adolescent offenders
• Developed specifically for probation officers and mental health professionals to administer
Youth Level of Service/Case Management Inventory (YLS/CMI)

Measure Description: 42 items measure 8 domains: (each item is coded as present or absent)

1. Prior and current offenses/dispositions
2. Family circumstances/parenting
3. Education/employment
4. Peer relations
5. Substance abuse
6. Leisure/recreation
7. Personality/behavior
8. Attitudes/orientation

Hoge & Andrews (2002)
Problem

Does the YLS/CMI possess sufficient predictive validity as it is used in the Nebraska Juvenile Justice System?
Basic Concepts

It’s all about error....
Basic Concepts

• Random Error
• Systematic Error
• Reliability
• Validity
Random Error

Unpredictable errors that go in different directions

- Fluctuations in measurement that are inconsistent in direction and magnitude
- Result from random individual differences in raters' emotions, attitudes, cognitive understanding
  
  - Temporal events that change over time in haphazard ways
  - Different people respond to the same stimulus materials in different ways that are unpredictable
Absence of random error

- Measurement that produces the same results repeatedly with the same stimulus materials
- Controls individual differences in raters’ emotions, attitudes, cognitive understanding as they impact behavior of interest
- Events are unchanged over time
- Different people respond to the same stimulus materials in the same predictable ways
Systematic Error

Predictable errors that go in same direction repeatedly

- Deviation in measurement that is consistent in direction and magnitude
- Result from fixed differences in types of individual respondents (e.g., personality or experience or biological differences)
  - Drift in measurement in one direction over time
  - People respond to an irrelevant component of complex stimulus materials in the same way regardless of the other relevant components
Validity

Absence of systematic error

• Measurement is consistent in direction and magnitude

• scores distribute around the true parameter

• Controls fixed differences in types of individual respondents (e.g., personality or experience or biological differences) as they influence the relevant behavior

• Absence of drift in measurement over time

• Control response to irrelevant components of complex stimulus materials
High Random Error – Low Systematic Error

Strike Zone, naive

Vertical Location (ft.)

Horizontal Location (ft.)
High Random Error – High Systematic Error

Strike Zone, naive

Vertical Location (ft.)

Horizontal Location (ft.)

X

X

X

X

X

X

X

X

X

X
Low Random Error – High Systematic Error

Strike Zone, naive

Vertical Location (ft.)

Horizontal Location (ft.)
Low Random Error – Low Systematic Error

Strike Zone, naive
Basic Concepts

How to tell if a measure is Valid?
Basic Concepts

- Predictive Validity
- Statistical Significance and Effect Sizes
- Main Effects and Interactions
- Logistic Regression
Basic Concepts

Predictive Validity

... is the extent to which an instrument predicts a criterion of interest

• Criterion for the YLS/CMI is failure in the juvenile justice system
Basic Concepts

Predictive Validity

• Higher levels of risk on the YLS/CMI should be associated with higher rates of failure.
• Lower levels of risk should be associated with lower rates of failure.
Basic Concepts

Effect Size

... is the strength of the relationship between the instrument and the criterion.

- The effect size for the YLS/CMI is the strength of the relationship between YLS scores and failure in the juvenile justice system.
Basic Concepts

Effect Sizes

… one common measure of effect size is the point-biserial correlation coefficient, “r”

• r ranges from -1.00 to 1.00

• Positive numbers indicate increases in risk are associated with increases of failure
Basic Concepts

Effect Sizes

\[ \text{... the value of } r \text{ between 0 and 1} \]
Indicates the strength of the effect size

- Small effect size: \( 0 < r \leq 0.10 \)
- Medium effect size: \( 0.10 < r \leq 0.35 \)
- Large effect size: \( 0.35 < r \leq 0.50 \)
- Very large effect size: \( 0.50 < r \leq 1.00 \)
Some Common Effect Sizes

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Dependent Variable</th>
<th>Sample Size</th>
<th>r as effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vietnam Veteran Status</td>
<td>Alcohol Problems</td>
<td>4,462</td>
<td>.44</td>
</tr>
<tr>
<td>Cigarette Smoking</td>
<td>Lung Cancer</td>
<td>1,385</td>
<td>.40</td>
</tr>
<tr>
<td>Psychotherapy</td>
<td>Mental Health</td>
<td>1,111</td>
<td>.38</td>
</tr>
<tr>
<td>Beta Carotene (Cancer Prevention)</td>
<td>Death</td>
<td>19,133</td>
<td>.20</td>
</tr>
</tbody>
</table>

Effect sizes at .20 and above show meaningful relationships.
Basic Concepts

Statistical Significance

... measures the probability of obtaining an effect size in a sample that is greater than 0 by chance alone.

• Convention: if $p < .05$ (5 out of 100), we accept it as statistically significant
Main Effect

... refers to the relationship between one predictor (here, the YLS/CMI score or level) and one outcome factor (here, failure in the criminal justice system).
Basic Concepts

Interaction Effect

... refers to *moderation* or the extent to which the effect of one variable depends upon the level of a second variable

**Example**: an interaction between sex of the youth and YLS score in predicting failure

- Does the YLS so a better job of predicting failure for boys than for girls? *(We hope not!)*
Basic Concepts

Logistic Regression

... predicts the outcome of a binary criterion (i.e., failure v. success) based on one or more predictor variables (e.g., possible time youth was in the system, YLS/CMI scores, gender and race).
Basic Concepts

Logistic Regression

... calculates the optimal weights for each predictor variable (Beta’s), effect sizes for each predictor variable (Odds ratios – converted to r’s) and tests the statistical significance of the predictors (with the Wald and Chi-square statistics).
Has anyone studied the validity of the YLS before?
The results of meta-analyses measure the strength of the relationship between the predictors (LSI criminogenic scales) and an outcome measure (recidivism) across multiple studies.

Olver et al. (2014) -- 128 studies of the LSI scales world wide:

YLS/CMI Effect sizes ($k = 36$ studies)
- Overall: $r = .25$
- Canada: $r = .33$
- Outside North America: $r = .28$
- United States: $r = .22$
YLS/CMI Validity Study in Nebraska!
Sample
(from Nebraska Probation)

• 6,158 individual juvenile probationers (one record per child), each of whom had an index YLS/CMI assessment (i.e., the first one within our time frame) between May 24, 2007 and November 11, 2015.
Youth’s Age at First YLS/CMI Assessment Date ($M = 15.5$ years old)
Gender of Youth Included in the Sample

- Male: 3948 (64.1%)
- Female: 2204 (35.8%)
Self-reported Race and Ethnicity Breakdown of the Youth Included in the Sample

- Caucasian European: 3207 (52.1%)
- African American: 1369 (22.3%)
- Hispanic: 1223 (19.9%)
- American Indian: 186 (3.0%)
- Asian/Pacific Islander: 58 (0.9%)
- Other: 109 (1.8%)
Distribution of YLS/CMI Scores for all the Youth Included in the Sample

Mean = 12.99
Std. Dev. = 6.048
N = 6,152
In Nebraska, total score places youth in one of four categories for future risk for continued criminal behavior:

- Low (0 to 8)
- Low Moderate (9 to 15)
- High Moderate (16 to 22)
- High (23 to 34)
- Very High (35 to 42) – there were only 2
YLS/CMI Total Risk Levels for all Youth Included in the Sample

- Low Risk: 1490 (53.5%)
- Low Moderate Risk: 2453 (22.4%)
- High Moderate Risk: 1715 (3.0%)
- High and Very High Risk: 455 (0.9%)
Validity Results: Main Effects

• **Success:** Youth with a successful first disposition and never returned to probation (success and no recidivism).

• **Failure:** Youth with an unsuccessful first disposition and/or returned to probation (unsuccessful and or recidivated).

Note: There are other ways to define success and failure (e.g., returned but with a successful disposition)
Predictive Validity of the YLS/CMI Total Score for Success Outcome – Predicting Failure (N=5782)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Beta</th>
<th>S.E.</th>
<th>Wald</th>
<th>d.f.</th>
<th>O.R.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Possible Time in System</td>
<td>.0001</td>
<td>.000</td>
<td>7.846*</td>
<td>1</td>
<td>1.000</td>
</tr>
<tr>
<td>YLS Total Score</td>
<td>.101</td>
<td>.005</td>
<td>430.223**</td>
<td>1</td>
<td>1.106</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.512</td>
<td>.080</td>
<td>358.578**</td>
<td>1</td>
<td>.220</td>
</tr>
</tbody>
</table>

Note: Model $\chi^2$ (2) = 478.527, $p < .001$; Nagelkerke $R^2 = .105$; $r = .29$; *$p < .01$. **$p < .001$.

$r = .29$

Remember meta-analysis $r = .22$ in the rest of the U.S.
Predictive Validity of the YLS/CMI Risk Level for Success Outcome – Predicting Failure (N=5782)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Beta</th>
<th>S.E.</th>
<th>Wald</th>
<th>d.f.</th>
<th>O.R.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Possible Time in System</td>
<td>.000</td>
<td>.000</td>
<td>5.572*</td>
<td>1</td>
<td>1.000</td>
</tr>
<tr>
<td>YLS/CMI Risk Level</td>
<td></td>
<td></td>
<td>369.802**</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Low Moderate vs. Low</td>
<td>.681</td>
<td>.071</td>
<td>90.933**</td>
<td>1</td>
<td>1.975</td>
</tr>
<tr>
<td>High Moderate vs. Low</td>
<td>1.338</td>
<td>.077</td>
<td>298.649**</td>
<td>1</td>
<td>3.810</td>
</tr>
<tr>
<td>High and Very High vs. Low</td>
<td>1.646</td>
<td>.122</td>
<td>183.215**</td>
<td>1</td>
<td>5.188</td>
</tr>
<tr>
<td>Constant</td>
<td>-.958</td>
<td>.069</td>
<td>191.379**</td>
<td>1</td>
<td>.384</td>
</tr>
</tbody>
</table>

Note: Model $\chi^2 (4)=395.860 \ p < .001$; Nagelkerke $R^2 = .088$; $r = .26$; *$p < .05$, **$p < .001$
Mean Probability of Failure at Each YLS/CMI Risk Level for the Success Outcome – Predicting Failure (N = 5782)

\[ \chi^2 (4) = 395.860, \ p < .001 \]

<table>
<thead>
<tr>
<th>YLS Level</th>
<th>Mean Probability</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>0.30</td>
<td>1435</td>
</tr>
<tr>
<td>low moderate</td>
<td>0.45</td>
<td>2329</td>
</tr>
<tr>
<td>high moderate</td>
<td>0.61</td>
<td>1613</td>
</tr>
<tr>
<td>high and very high</td>
<td>0.68</td>
<td>405</td>
</tr>
</tbody>
</table>

Note: Means sharing letters are not different
Validity Results: Interactions

- **Moderation**: Does the predictive validity of the YLS/CMI in Nebraska vary by gender or by minority-majority status?
Effects of Gender, YLS/CMI Total Risk Score and their Interaction on Success Outcome – Predicting Failure (N = 5782)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Beta</th>
<th>S.E.</th>
<th>Wald</th>
<th>d.f.</th>
<th>O.R.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Possible Time in System</td>
<td>.000</td>
<td>.000</td>
<td>8.086**</td>
<td>1</td>
<td>1.000</td>
</tr>
<tr>
<td>YLS Total Score</td>
<td>.099</td>
<td>.006</td>
<td>270.800**</td>
<td>1</td>
<td>1.104</td>
</tr>
<tr>
<td>Gender</td>
<td>-.331</td>
<td>.144</td>
<td>5.282*</td>
<td>1</td>
<td>.718</td>
</tr>
<tr>
<td>Gender * YLS Total Score</td>
<td>.004</td>
<td>.010</td>
<td>.168</td>
<td>1</td>
<td>1.004</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.393</td>
<td>.094</td>
<td>218.501**</td>
<td>1</td>
<td>.248</td>
</tr>
</tbody>
</table>

Note: Model $\chi^2 (4) = 502.114, p < .001$; Nagelkerke $R^2 = .110; r = .005; *p < .05. **p < .01.$
Mean Probability of Failure at Each YLS/CMI Risk Level for Boys (N = 3711) and Girls (N = 2071)

Boys

<table>
<thead>
<tr>
<th>YLS Level</th>
<th>Mean Predicted Probability</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>0.30</td>
<td>1435</td>
</tr>
<tr>
<td>low moderate</td>
<td>0.45</td>
<td>2329</td>
</tr>
<tr>
<td>high moderate</td>
<td>0.61</td>
<td>1613</td>
</tr>
<tr>
<td>high and very high</td>
<td>0.68</td>
<td>405</td>
</tr>
</tbody>
</table>

χ²(4) = 395.860, p < .001

Note: Means sharing letters are not different

Girls

<table>
<thead>
<tr>
<th>YLS Level</th>
<th>Mean Predicted Probability</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>0.26</td>
<td>551</td>
</tr>
<tr>
<td>low moderate</td>
<td>0.41</td>
<td>834</td>
</tr>
<tr>
<td>high moderate</td>
<td>0.56</td>
<td>559</td>
</tr>
<tr>
<td>high and very high</td>
<td>0.65</td>
<td>127</td>
</tr>
</tbody>
</table>

χ²(8) = 419.928, p < .001

Note: Means sharing letters are not different
## Effects of Race/Ethnicity, YLS/CMI Risk Score and their Interactions on Failure (N = 5280)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Beta</th>
<th>S.E.</th>
<th>Wald</th>
<th>d.f.</th>
<th>O.R.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Possible Time in System</td>
<td>.000</td>
<td>.000</td>
<td>6.989**</td>
<td>1</td>
<td>1.000</td>
</tr>
<tr>
<td>YLS Total Score</td>
<td>.100</td>
<td>.007</td>
<td>225.692***</td>
<td>1</td>
<td>1.105</td>
</tr>
<tr>
<td>Minority Status</td>
<td></td>
<td></td>
<td>3.869ns</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>White v. Black</td>
<td>.336</td>
<td>.176</td>
<td>3.660ns</td>
<td>1</td>
<td>1.400</td>
</tr>
<tr>
<td>White v. Hispanic</td>
<td>.013</td>
<td>.190</td>
<td>.005ns</td>
<td>1</td>
<td>1.013</td>
</tr>
<tr>
<td>Minority Status * YLS Total Score</td>
<td></td>
<td></td>
<td>1.294ns</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>White v. Black * YLS Total Score</td>
<td>-.001</td>
<td>.012</td>
<td>.012ns</td>
<td>1</td>
<td>.999</td>
</tr>
<tr>
<td>White v. Hispanic * YLS Total Score</td>
<td>.014</td>
<td>.014</td>
<td>1.136ns</td>
<td>1</td>
<td>1.015</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.617</td>
<td>.103</td>
<td>244.561***</td>
<td>1</td>
<td>.198</td>
</tr>
</tbody>
</table>

Note: Model $\chi^2$ (6) =476.833, $p < .001$; Nagelkerke $R^2 = .113$; *$p < .05$. **$p < .01$, $p < .001$. 
Mean Probability of Failure at Each YLS/CMI Risk Level for European Caucasian Youth (N = 3028)

$\chi^2 (3) = 215.319, p < .001; \text{Nagelkerke } R^2 = .092$

<table>
<thead>
<tr>
<th>YLS Level</th>
<th>Mean Predicted Probability</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>0.26</td>
<td>788</td>
</tr>
<tr>
<td>low moderate</td>
<td>0.43</td>
<td>1183</td>
</tr>
<tr>
<td>high moderate</td>
<td>0.58</td>
<td>846</td>
</tr>
<tr>
<td>high and very high</td>
<td>0.65</td>
<td>211</td>
</tr>
</tbody>
</table>

Note: Means sharing letters are not different
Mean Probability of Failure at Each YLS/CMI Risk Level for African American Youth (N = 1280)

\[ \chi^2 (2) = 75.594, p < .001; \text{Nagelkerke } R^2 = .077 \]

<table>
<thead>
<tr>
<th>YLS Level</th>
<th>Mean Predicted Probability</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>0.35</td>
<td>286</td>
</tr>
<tr>
<td>low moderate</td>
<td>0.50</td>
<td>526</td>
</tr>
<tr>
<td>high moderate</td>
<td>0.66</td>
<td>386</td>
</tr>
<tr>
<td>high and very high</td>
<td>0.70</td>
<td>82</td>
</tr>
</tbody>
</table>

Note: Means sharing letters are not different
Mean Probability of Failure at Each YLS/CMI Risk Level for Hispanic Youth (N = 3028)

\(\chi^2 (2) = 87.357, p < .001; \text{ Nagelkerke } R^2 = .104\)

<table>
<thead>
<tr>
<th>YLS Level</th>
<th>Mean Predicted Probability</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>0.33</td>
<td>262</td>
</tr>
<tr>
<td>low moderate</td>
<td>0.44</td>
<td>414</td>
</tr>
<tr>
<td>high moderate</td>
<td>0.66</td>
<td>291</td>
</tr>
<tr>
<td>high and very high</td>
<td>0.73</td>
<td>75</td>
</tr>
</tbody>
</table>

Note: Means sharing letters are not different
Conclusions:

1. The YLS/CMI demonstrates validity with respect to predicting failure in the juvenile justice system.
   - Continuous scale $r = .29$
   - Risk levels show significant step function in expected direction
Conclusions:

2. The YLS/CMI shows no evidence of disparate impact in the way it predicts failure by gender.
   • Boys are significantly higher on failure than girls
Conclusions:

3. The YLS/CMI shows no evidence of disparate impact in the way it predicts failure by minority status.

4. Using risk level and not risk score as a predictor
   - African Americans and Hispanics are somewhat more likely to fail than are European Americans
Thank you for your time and patience!
Effects of Race/Ethnicity, YLS/CMI Risk Level and their Interactions on Failure (N = 5280)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Beta</th>
<th>S.E.</th>
<th>Wald</th>
<th>d.f.</th>
<th>O.R.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Possible Time in System</td>
<td>.000</td>
<td>.000</td>
<td>4.820*</td>
<td>1</td>
<td>1.000</td>
</tr>
<tr>
<td>YLS/CMI Risk Level</td>
<td></td>
<td></td>
<td>200.850***</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Low Moderate vs. Low</td>
<td>.756</td>
<td>.100</td>
<td>56.930***</td>
<td>1</td>
<td>2.130</td>
</tr>
<tr>
<td>High Moderate vs. Low</td>
<td>1.370</td>
<td>.107</td>
<td>163.602***</td>
<td>1</td>
<td>3.934</td>
</tr>
<tr>
<td>High and Very High vs. Low</td>
<td>1.691</td>
<td>.166</td>
<td>103.637***</td>
<td>1</td>
<td>5.427</td>
</tr>
<tr>
<td>Minority status</td>
<td></td>
<td></td>
<td>9.491**</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>White vs. Black</td>
<td>.424</td>
<td>.148</td>
<td>7.802*</td>
<td>1</td>
<td>1.513</td>
</tr>
<tr>
<td>White vs. Hispanic</td>
<td>.317</td>
<td>.155</td>
<td>4.190*</td>
<td>1</td>
<td>1.372</td>
</tr>
<tr>
<td>Minority status * YLS/CMI Risk Level</td>
<td></td>
<td></td>
<td>4.493ns</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>White vs. Black * YLS * Low Moderate vs. Low</td>
<td>-.145</td>
<td>.182</td>
<td>.634ns</td>
<td>1</td>
<td>.865</td>
</tr>
<tr>
<td>White vs. Black * YLS * High Moderate vs. Low</td>
<td>-.094</td>
<td>.196</td>
<td>.229ns</td>
<td>1</td>
<td>.911</td>
</tr>
<tr>
<td>White vs. Black * YLS * High/V. High vs. Low</td>
<td>-.231</td>
<td>.317</td>
<td>.532ns</td>
<td>1</td>
<td>.794</td>
</tr>
<tr>
<td>White vs. Hispanic * YLS * Low Moderate vs. Low</td>
<td>-.269</td>
<td>.191</td>
<td>1.988ns</td>
<td>1</td>
<td>.764</td>
</tr>
<tr>
<td>White vs. Hispanic * YLS * High Moderate vs. Low</td>
<td>.031</td>
<td>.210</td>
<td>.022ns</td>
<td>1</td>
<td>1.032</td>
</tr>
<tr>
<td>White vs. Hispanic * YLS * High/V. High vs. Low</td>
<td>.055</td>
<td>.336</td>
<td>.027ns</td>
<td>1</td>
<td>1.056</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.123</td>
<td>.090</td>
<td>155.943***</td>
<td>1</td>
<td>.325</td>
</tr>
</tbody>
</table>

Note: Model $\chi^2$ (12) =404.632, $p < .001$; Nagelkerke $R^2 = .097$; $r = .26$; *$p < .05$, **$p < .01$, ***$p < .001$