Developing, Implementing, and Evaluating a Screening Assessment for Maryland Social Services Administration

A joint initiative of Maryland Department of Human Resources, Social Services Administration; Casey Family Programs; and the Children's Research Center (CRC)

Deirdre O'Connor, LCSW; CRC
Debbie Ramelmeier, LCSW-C, J.D.; Maryland SSA
Many thanks to Maryland Social Services Agency for initiating and completing this work and Casey Family Programs for supporting this project.
Agenda

- Impetus for development
- Assessment development
- Pilot implementation
- Statewide implementation and evaluation
  - Methods
  - Findings
- Supporting implementation
- Answering your questions
Impetus for Development

- County managers raised concerns with state administrators
- Anecdotal evidence that there was inconsistency in screening decisions between jurisdictions
- Large disparity in screening rates
- Screening decision relied on local interpretation of state policy
Incorporating Research Into Practice

- Maryland SSA administrators recognized need for improvement, not just change
- Pilot implementation and evaluation were needed prior to statewide implementation
- Substantial evaluation activities were always part of implementation plan
Screening Assessment Development

• Based on Maryland law, policy, and regulation

• CRC staff facilitated several meetings with local agency and state office staff

• Great deal of time needed to refine and clarify policy; local policy interpretation evident

• Developed structure and definitions for screening and response time tool
Screening and Response Time Assessment Pilot

- Assessment development: Spring 2008
- Assessment pilot: July 2008
  - Baltimore City, Montgomery County, and Anne Arundel County
  - Training focused on screening tool structure and definitions
  - Screening tool completed outside of SACWIS
- Evaluation of pilot: October 2008
  - Pre- and post-implementation case file review
  - Initial reliability test
Statewide Implementation

• Pilot evaluation identified areas for improvement
  » Clarified several definitions
  » Expanded training to include narrative documentation

• Statewide training: January 2009
  » Explicitly stated goal of increased consistency
  » Included description of post-implementation evaluation activities

• Statewide implementation: February 2009 (still documented outside of SACWIS)
<table>
<thead>
<tr>
<th>Research Question</th>
<th>Evaluation Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does the assessment help workers make more consistent screening decisions?</td>
<td>Inter-rater reliability testing</td>
</tr>
<tr>
<td>Has it influenced screening practices?</td>
<td>Qualitative case review</td>
</tr>
<tr>
<td>Are workers writing more precise narrative?</td>
<td></td>
</tr>
<tr>
<td>Are workers completing the assessment as intended?</td>
<td>Survey of workers</td>
</tr>
<tr>
<td>Are they completing it prior to making the decision?</td>
<td></td>
</tr>
</tbody>
</table>
Testing the Assessment’s Reliability: Inter-rater Agreement on Case Vignettes

Description

• Forty-six workers from 22 jurisdictions

• Thirty-six referral vignettes were drawn from actual records in CHESSIE

• Each worker completed the screening assessment on 12 vignettes

Measures

• Rate of agreement for screening decision and items

• Kappa statistic
## Testing the Assessment’s Reliability: Inter-rater Percent Agreement Findings

<table>
<thead>
<tr>
<th>Item Examined</th>
<th>Average Rate of Agreement</th>
<th>Minimum Rate of Agreement</th>
<th>Maximum Rate of Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial decision</td>
<td>87.9%</td>
<td>53.8%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Final decision after overrides</td>
<td>87.6%</td>
<td>50.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Inter-rater agreement across individual items</td>
<td>89.5 – 99.8%</td>
<td>50.0 – 94.4%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>
Testing the Assessment’s Reliability: Fleiss’ Kappa Findings

<table>
<thead>
<tr>
<th>Item Examined</th>
<th>Average Fleiss’ Kappa Across 36 Cases (Confidence Interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability for the 28 items and decision across intake workers</td>
<td>.64 (.61–.68)</td>
</tr>
<tr>
<td>Reliability for maltreatment classifications and decision across intake workers</td>
<td>.76 (.68–.84)</td>
</tr>
</tbody>
</table>
Assessment Reliability Findings: Summary

- High rates of agreement among workers who voluntarily participated in testing
  - Percent agreement for screening decision was 75% or better for 32 of 36 vignettes
  - Agreement rate was 90% or higher for each of the 28 assessment items
- Fleiss’ kappa similar to those of other screening assessments
- Findings suggest the screening and response time assessment and its associated item definitions can help workers make more consistent screening decisions.
Case File Review: Description of Method

Pre-implementation case review:

- Provided a baseline measure of documentation quality
- 196 randomly selected reports
  - Non-pilot agencies
  - September 2008

Post-implementation case review:

- Focused on accuracy of completed screening assessments relative to narrative and other case file documentation
- Quality of documentation
- 244 randomly selected reports
  - Pilot and non-pilot agencies
  - April 2009
## Case File Review: Pre- and Post-implementation Comparison

<table>
<thead>
<tr>
<th>Item Examined</th>
<th>Pre-implementation Case Reading</th>
<th>Post-implementation Case Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Reports Reviewed</td>
<td>164</td>
<td>244</td>
</tr>
<tr>
<td>Screening decision supported by narrative</td>
<td>93.3%</td>
<td>95.9%</td>
</tr>
<tr>
<td>Allegation types</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. All supported by narrative*</td>
<td>65.8%</td>
<td>73.8%</td>
</tr>
<tr>
<td>b. Some supported by narrative</td>
<td>11.0%</td>
<td>9.0%</td>
</tr>
<tr>
<td>c. None supported by narrative</td>
<td>23.2%</td>
<td>17.2%</td>
</tr>
<tr>
<td>Allegation not marked that should have been*</td>
<td>23.8%</td>
<td>7.3%</td>
</tr>
<tr>
<td>Response time supported by narrative*</td>
<td>74.4%</td>
<td>84.2%</td>
</tr>
</tbody>
</table>

*Indicates significant difference in rates (z score, p < .05, two-tailed test).
Case File Review: Summary of Findings

- Post-implementation, significantly more reports had:
  - Narratives that fully documented all the maltreatment allegations indicated in the report
  - Narratives that matched allegations recorded in CHESSIE
  - Narratives that justified the response time assigned
- Inconsistent interpretation of some neglect allegation definitions
- Implementation fidelity varied by office
- Overall, the screening assessment and narrative training efforts improved workers’ case narratives.
Worker Survey: Description of Method

- Web-based survey of intake staff in June 2009
- 39.7% response rate (73 of 184)
- Data collected:
  - Structured questions (yes/no and Likert scale)
  - Open-ended questions
  - Basic demographics: age, experience, and degree
- Content:
  - Ease of completion
  - Perceived usefulness
  - Effect on screening decisions and practices
Worker Survey: Self-reported Use of Assessment

(N = 73 workers, June 2009)

- Have referred to definitions when completing an assessment: 57.5%
- Have applied an override to the assessment: 61.6%
- Have referenced the assessment while following up on a report: 13.7%
## Worker Survey: Self-reported Frequency of Assessment Completion

<table>
<thead>
<tr>
<th>Activity</th>
<th>Average</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definitions were referenced during completion</td>
<td>25%</td>
<td>1–100%</td>
</tr>
<tr>
<td>Applied overrides to screening assessment</td>
<td>9%</td>
<td>1–50%</td>
</tr>
<tr>
<td>Discussed an override with supervisor</td>
<td>8%</td>
<td>1–100%</td>
</tr>
</tbody>
</table>

Note: Unit = Percentage of reports for which activity occurs, on average.
Worker Survey: Summary of Findings

- Most workers (75–80%) knew the screening assessment was intended to inform the screening decision and reflect agency policy.

- Two thirds (57–61%) of workers referenced assessment definitions during completion.

- Approximately 60% of respondents found the assessment helpful.

- Practice appears to vary by worker and sometimes by office.
<table>
<thead>
<tr>
<th>Question</th>
<th>Method</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do the screening and response time assessment and definitions help workers obtain consistent screening decisions?</td>
<td>Inter-rater reliability testing</td>
<td>Yes</td>
</tr>
<tr>
<td>Has using the screening and response time assessment positively influenced other aspects of screening?</td>
<td>Qualitative case review</td>
<td>Yes</td>
</tr>
<tr>
<td>Are workers completing the screening and response time assessment as intended and prior to making the decision?</td>
<td>Survey of workers</td>
<td>Sometimes</td>
</tr>
</tbody>
</table>
Implications of the Research

- SSA’s project used research to inform practice in the following two ways:
  1. Assessment development
  2. Evaluation of implementation fidelity

- The project demonstrates to other agencies:
  » How structured assessments can help improve decision making
  » How research can support assessment development and implementation
State office can produce reports on screened-in/-out rates and the frequency of “overrides.” This type of monitoring:

- Allows for targeted investigation and training when numbers show wider than acceptable variance between local department screening activity
- Points to areas where more work is needed to refine policy and definitions
- Provides information so local departments can evaluate their screening rates against their sister departments
Supporting Implementation

• Incorporated the assessment into SACWIS

• Supporting implementation
  » Monitoring and reporting on assessment implementation
  » Incorporating assessment findings into case file reviews to increase implementation fidelity
Contact Information

Deirdre O’Connor
NCCD-CRC
301-920-0638
doconnor@nccdcrc.org

Debbie Ramelmeier
Maryland SSA
410-767-7506
DRamelme@dhr.state.md.us