

**Implementation and Evaluation  
of Maryland Social Services Administration's  
Screening and Response Time Assessment**

*Kristen Johnson  
Deirdre O'Connor*

*October 2009*



---

*A nonprofit social research agency and division of the National Council on Crime and Delinquency*  
426 South Yellowstone Drive, Suite 250  
Madison, WI 53719  
608-831-1180  
[www.nccd-crc.org](http://www.nccd-crc.org)

## TABLE OF CONTENTS

EXECUTIVE SUMMARY .....	i
I. INTRODUCTION .....	1
A. Background .....	1
B. Description of Screening and Response Time Assessment Development and Implementation .....	3
1. Field Test of the Screening and Response Time Assessment.....	3
2. Evaluation of Statewide Implementation.....	5
II. TESTING INTER-RATER RELIABILITY .....	6
A. Methodology .....	6
B. Findings.....	8
C. Summary .....	11
III. EXAMINING PRACTICE WITH PRE- AND POST-IMPLEMENTATION CASE REVIEWS.....	12
A. Methodology .....	12
B. Findings.....	13
C. Summary .....	15
IV. SOLICITING THE INPUT OF SCREENING STAFF .....	16
A. Methodology .....	16
B. Findings.....	16
C. Summary .....	21
V. DISCUSSION .....	21
A. Summary of Findings.....	21
B. Implications.....	23
REFERENCES .....	25

### APPENDICES

Appendix A: Additional Inter-rater Reliability Findings

Appendix B: Web-based Survey of Workers

Appendix C: SDM<sup>®</sup> Child Abuse and Neglect Screening and Response Time Assessment Tool and Definitions

Appendix D: Findings From the Field Test

Children’s Research Center is a nonprofit social research organization and  
a division of the National Council on Crime and Delinquency.

Structured Decision Making<sup>®</sup> and SDM<sup>®</sup> are  
registered in the U.S. Patent and Trademark Office.

## **ACKNOWLEDGMENTS**

This study was a joint effort involving the Children’s Research Center (CRC), the Maryland Department of Human Resources Social Service Administration (SSA), and Casey Family Programs. CRC would like to thank Casey Family Programs staff Peter Pecora and Oronde Miller; SSA administrators Deborah Ramelmeier and Steve Berry, and former SSA administrators Cathy Mols and Heather Stowe for their thoughtful collaboration in the development, field test, and statewide implementation and evaluation of the screening and response time assessment. Thank you to the administrators and staff of Anne Arundel County, Baltimore City, and Montgomery County for their participation in the development and field-testing of the assessment. Finally, this study would not have been possible without the commitment and cooperation of the staff in all the local departments of social services for the state of Maryland who participated in the reliability test and worker survey.

## EXECUTIVE SUMMARY

In an effort to ensure reliable decision making in child protective services (CPS) screening, administrators of Maryland’s Social Services Administration (SSA), in collaboration with Casey Family Programs, contracted with the Children’s Research Center (CRC), a nonprofit social research organization and division of the National Council on Crime and Delinquency, to 1) Develop a structured screening and response time assessment; 2) implement the assessment statewide; and 3) evaluate the impact of implementation on screening decisions. The purpose of evaluating statewide implementation was to determine whether the screening and response time assessment helped workers reach reliable decisions given the same information, and whether implementation improved narrative support workers provided for the screening and response time decisions made. Evaluation activities included reliability testing of the assessment, a qualitative review of screening decisions conducted before and after implementation of the assessment, and a survey of workers about the assessment and its implementation. Table E1 reviews the pertinent questions, methods for determination, and the answer found during the evaluation.

<b>Table E1</b>		
<b>Research Question by Evaluative Activity</b>		
<b>Question</b>	<b>Method</b>	<b>Answer</b>
Do the screening and response time assessment and its associated definitions help workers obtain consistent screening decisions when provided with the same information?	Inter-rater reliability testing	Yes
Has using the screening and response time assessment to structure the decision influenced practices involved in screening a report?  Are workers writing more precise narrative?	Qualitative case review	Yes
Are workers completing the screening and response time assessment as intended according to policy and best practice standards?  Are they completing it prior to making the screening decision?	Survey of workers	Sometimes

The evaluation showed that the screening and response time assessment can be useful to workers. The test of inter-rater reliability among field staff demonstrated that the assessment resulted in significant levels of agreement. The rate of agreement for the screening decision was 75% or greater for all but four cases, and averaged 88% overall.

Findings from the pre- and post-implementation case readings suggest that implementing the screening and response time assessment, including training on report narration, improved workers’ documentation within the case narrative and possibly improved the screening decisions they made. In the post-implementation case reading, significantly more reports had narratives that fully justified all types of maltreatment indicated in the report, and also had narrative justifying the response time.

A web-based survey of screening staff confirmed that the assessment can aid decision making. Some respondents noted, however, that the quality of implementation varied across workers and offices. More than half (60%) of the 73 workers who responded to the web-based solicitation of feedback found the assessment helpful and had referenced the definitions at least once. Workers

who found the assessment helpful noted that it clarified policy, provided a decision framework for new workers, and helped when evaluating more difficult reports. Workers who did not find the assessment useful indicated that it is additional documentation, and they often complete the assessment after making a decision about the referral. SSA administrators plan to add the screening and response time assessment to MD CHESSIE, which will require that screening workers complete it prior to making the screening decision.

These findings strongly suggest that workers can increase the consistency of decisions whether or not to screen in a maltreatment report with use of the screening and response time assessment, but this requires consistent implementation and use of definitions. Research demonstrates that good implementation involves strong administrative support for the practice change, effective training and supervision, and skilled staff (Toth, Manly, & Nilsen, 2008; Mihalic, Irwin, Fagan, Ballard, & Elliott, 2004). Maryland Department of Human Resources Social Services Administration (SSA) administrators may wish to strengthen implementation in some offices by employing efforts used by other jurisdictions, such as the following:

- Emphasize staff use of definitions to promote accurate and consistent assessment scoring. Ensuring that definitions are easily accessible to workers may increase the consistency of decisions.
- Encourage staff and supervisors to reference the screening assessment when reviewing reports. For example, Michigan's Department of Human Services developed a comparative case reading program designed to improve supervisors' evaluation of workers' decision and documentation practices and to improve workers' assessment practices. Supervisors review a sample of case files, and then quality experts review the same file. The supervisors' findings can then be compared to the experts' findings. Findings are reviewed in a summary meeting with supervisors and area managers.
- Solicit staff questions and identify areas for follow-up training or additional emphasis. If clarification of policy or intent is needed, staff may want to respond with a written question-and-answer list, ask supervisors to review the subject at a future staff meeting, or revise training materials to include a case example that addresses the issue.
- Continue efforts to improve practice, such as conducting cross-office qualitative case reviews, communicating program successes to staff, and engaging managers and staff in discussions about how to further increase the validity and consistency of decision making.
- Encourage supervisors to routinely review assessment completion and incorporate it into case discussions with workers.
- Ensure that assessment and case action data for child protective services (CPS) reports are easily accessible to staff.

One of the most effective strategies for improving practice identified through the Child and Family Services Review (CFSR) is monitoring practice with data (Administration on Children and Families, 2006). Regular examination of assessment findings through data reporting can determine how often workers are completing the assessment, what evidence is provided for the screening decision, and if assessment findings are consistent with the final screening decision. If managers had access to such information, they could better monitor and therefore strengthen field practice.

## **I. INTRODUCTION**

In an effort to ensure reliable decision making in child protective services (CPS) screening, administrators of Maryland's Social Services Administration (SSA), in collaboration with Casey Family Programs, contracted with the Children's Research Center (CRC), a nonprofit social research organization and division of the National Council on Crime and Delinquency, to 1) Develop a structured screening and response time assessment; 2) implement the assessment statewide; and 3) evaluate the impact of implementation on screening decisions. This report describes the construction, implementation, and evaluation of the screening and response time assessment.

### **A. Background**

CPS agencies receive thousands of child maltreatment reports each year. In each case, a screening worker must decide in a relatively short timeframe which reports necessitate an in-person assessment, often with limited information from anonymous reporters. The in-person assessments require that a worker investigate the allegations to ensure the safety of a child, and involve considerable staff time and effort. These are difficult but critical decisions. Maryland's SSA sought to increase the consistency with which this decision is made by developing and implementing a screening assessment that is based on Code of Maryland regulations (COMAR) and SSA policy.

The decisions a screening worker makes (e.g., whether to screen in a report for further investigation) can influence the course of an individual's life, and implementing simple assessments can increase the reliability of such decisions. Research shows that case decisions based on clinical judgment alone have little predictive validity (Andrews, Bonta, & Wormith, 2006) and are unreliable among even the most qualified and experienced workers. For example,

Chapin Hall conducted inter-rater reliability testing for key decisions made by CPS workers regarding family services and child removals from homes (Rossi, Schuerman, & Budde, 1996). The study compared the inter-rater reliability of case recommendations among 27 child welfare experts (nationally recognized practitioners) for 70 case vignettes. The vignettes ranged in length from three to five pages, and each was based on an actual case. In addition, the researchers tested inter-rater reliability for case decisions among 103 workers from three different states (each rated 18 cases). Both the child welfare experts and the workers had low rates of agreement when deciding whether or not to place a child in foster care (agreement was 65% among experts and 64% among workers, with corresponding kappas of .45 and .35). Additional testing added in-home services as a third option, but this lowered the rate of agreement.

These findings have been replicated in other human services fields. One study examined inter-rater reliability among 82 staff when deciding whether to admit someone to a community-based mental health center (n = two cases; Hendryx & Rohland, 1997). Staff varied widely in their treatment recommendations, although all were very confident in the accuracy of their decision making. Prior studies reported similar findings (Munro, 1998).

Why is the reliability of workers' screening decisions important? A state agency wants to ensure that the decision whether or not to investigate a child maltreatment report is made according to policy and that policy is applied consistently between workers and across local offices. The implication of unreliable screening assessments is that the decision whether to recommend in-person investigation is based on worker discretion rather than state policy, making it difficult to achieve equity.

The consistency and validity of workers' decisions to recommend a report be investigated may increase with the use of a screening assessment that clarifies policy in a concise and user-



friendly format. Evidence from CPS suggests that completing simple, relatively objective assessments results in more consistent and accurate decision making (Baird, Wagner, Healy, & Johnson, 1999; Baird & Wagner, 2000). The next section describes how a screening and response priority assessment was developed for Maryland screening workers.

## **B. Description of Screening and Response Time Assessment Development and Implementation**

The initial screening and response time assessment was developed by a workgroup of agency staff and CRC consultants. The task of the workgroup was to translate existing policy and procedures, both legislative and departmental, into a concise list of observations that workers could reference to determine whether a report should be screened in for further investigation. The resulting screening assessment listed by maltreatment type the nature of reported allegations that workers should consistently screen in for further investigation. Each criterion for screening in a report was supplemented with a full definition to help guide workers in their selection (please refer to Appendix C to review the final assessment).

### **1. Field Test of the Screening and Response Time Assessment**

Once an initial assessment had been drafted and approved by agency managers, workers in three jurisdictions piloted the assessment in July 2008. This field test enabled workers to test the clarity of definitions and identify situations in which worker judgment and past practice differed from the guidelines represented by the screening assessment. In addition, the field test provided an opportunity to determine how well the screening assessment helped workers reach reliable decisions and provide narrative support for those decisions. The field test included two evaluation components:

- Inter-rater reliability testing, which showed that completing the screening assessment resulted in high rates of agreement among workers presented with the same case information. To test inter-rater reliability, managers selected 16 workers to complete the piloted assessment for 12 vignettes. Participating workers had 92% agreement for the final screening decision and over 80% agreement for individual items on the screening assessment.
- A pre- and post-implementation review of cases, which showed that the screening assessment improved the quality of case narratives and justifications for both the screening and the response priority decisions (see Appendix D for the full report).

The field test resulted in improvements both to the assessment and the process of implementation. As a result of these findings, agency managers further clarified items such as risk of harm and inadequate supervision. In addition, they changed the item “caregiver action that is likely to cause injury” to “caregiver action that likely caused injury.” Such changes are likely to increase the consistency of scoring by workers.

The field test also showed that some workers provided insufficient documentation for their decision making. Prior to implementing the screening and response priority assessment, 30% of the cases reviewed lacked sufficient narrative to justify allegations recorded, and 45% lacked narrative supporting the assigned response time. Implementing the assessment had a positive effect on case narratives. Among the 60 reports reviewed post-implementation, 20% lacked narrative justifying the allegations recorded, and 24% lacked narrative justifying the response time. In an attempt to further improve narrative quality, a targeted training about how to document a report was added to the statewide implementation plan. The field test showed that the screening assessment, when accompanied by strategies to ensure successful implementation, can improve the consistency of screening decisions made as well as narrative justification for the decisions. It also resulted in stronger assessment items and a more comprehensive approach to training.

2. Evaluation of Statewide Implementation

The purpose of evaluating statewide implementation was to determine whether the screening and response time assessment helped workers reach reliable decisions, and whether implementation improved the narrative support workers provided for screening decisions they made. Evaluation activities included reliability testing of the assessment, a qualitative review of screening decisions conducted before and after implementation, and a survey of workers about the assessment and its implementation. Specifically, the evaluation sought to answer the following questions (see Table 1):

1. Do the screening and response time assessment and its associated definitions help workers obtain consistent screening decisions when provided with the same type and depth of information?
2. Has using the screening and response time assessment to structure the decision influenced practices involved in screening a report?
3. Are workers writing more precise narrative?
4. Are workers completing the screening and response time assessment as intended by policy and best practice recommendations?
5. Are workers completing the screening and response time assessment prior to making the decision?

<b>Table 1</b>	
<b>Research Question by Evaluative Activity</b>	
Do the screening and response time assessment and its associated definitions help workers obtain consistent screening decisions when provided with the same information?	Inter-rater reliability testing using actual case narratives
Has using the screening and response time assessment to structure the decision influenced practices involved in screening a report?  Are workers writing more precise narrative?	Qualitative case review
Are workers completing the screening and response time assessment as intended according to policy and best practice standards?  Are they completing it prior to making the screening decision?	Survey of workers

The remainder of the report reviews the methods and findings from this research. The first section describes how the screening assessment's inter-rater reliability was tested; the second summarizes findings from qualitative report reviews conducted before and after screening implementation. The third section highlights feedback from surveyed workers about the screening assessment and its implementation.

## **II. TESTING INTER-RATER RELIABILITY**

To assess whether the screening assessment and its associated definitions helped workers reach the same decision given the same information, a sample of workers from across Maryland participated in inter-rater reliability testing. This test occurred in March 2009, two months after the assessment was implemented statewide and one month after a statewide training on how to document a report and screening decision.

### **A. Methodology**

Directors selected two or three workers with varying degrees of experience from each office to participate in inter-rater reliability testing. Approximately 60 workers were invited to participate in one of three testing sessions, each taking approximately two to three hours (46 workers actually participated). Participants were briefed on test procedures that emphasized the use of definitions while completing the vignettes; they then completed assessments for 12 of 36 prepared vignettes. Case vignettes consisted of actual reports made to Maryland DSS (stripped of identifying information), and varied in length between 100 and 300 words.<sup>1</sup> The objective of the test was to examine the consistency of workers' responses based on screening item definitions applied to the same case vignettes. Participants completed the screening

---

<sup>1</sup> Case vignettes were created by pulling the complete case narrative from referrals made to Maryland DSS and entering it into CHESSIE. Identifying information was removed from each vignette, and in some cases, client characteristics that would not impact the screening decision were changed to further disguise the referral.

assessment using the web-based screening assessment, which includes immediate access to definitions for each item.

CRC staff initially analyzed the results of the inter-rater reliability test using percent agreement, which is the percentage of times raters reached the same conclusion for each item on the assessment. Percent agreement findings are easy to understand and enable straightforward comparisons. Their only limitation is that percent agreement does not control for the degree to which different raters might select the same response by chance. Gwets (2008) defines chance agreement as that which “occurs when at least one rater rates an individual randomly” (Gwets, 2008, p. 35). One might speculate that chance has little influence on whether individuals agree on case actions in a child welfare context. It seems unlikely that workers make case decisions randomly; it is more likely that when stressed or working under difficult conditions, they will rely on prior cognitive maps to inform decision making (Campbell, Whitehead, & Finkelstein, 2009). It is true, however, that percent agreement does not correct for the likelihood that two individuals randomly select the same option for the same case. In addition, social workers completing vignettes in a testing situation might be more likely to select an answer at random.

Existing statistics to assess the reliability of categorical measures after controlling for chance have limitations. Cohen’s kappa, the most widely known measure of categorical reliability, is for comparisons between only two raters (Cohen, 1960). Fleiss’ kappa is the recommended statistic for categorical measures and more than two raters (Landis & Koch, 1977a). But when the traits being measured are rare or very frequent, Fleiss’ kappa yields low to negative inter-rater reliability statistics (Gwets, 2008), even in the presence of a true high rate of agreement (Rodella, 1995). Other techniques for assessing the reliability of categorical measures require modeling and thus a large number of cases. Workers would need to score a large number of vignettes, which is not feasible given current workload conditions. Despite its limitations,

Fleiss' kappa is the best statistical test for determining whether levels of agreement exceed chance for this sample of vignettes. The kappa statistic is reported for screening assessment items but should be considered descriptive only since assessment item frequencies will tend to be either high or low (for example, a neglect report is unlikely to have physical or sexual abuse items scored).

## **B. Findings**

Table 2 shows the average, minimum, and maximum rates of agreement for individual assessment items as well as the overall decision. Participating screening workers averaged 88% agreement for the final screening decision. Although the average percent agreement was 88%, agreement was lower (between 50 and 69%) for four of the 36 vignettes. These four reports involved neglect situations (for more details, see Appendix A), and discussing these vignettes with staff may increase the consistency of screening decisions for similar neglect reports. Among the remaining 32 vignettes, percent agreement was 75% or higher, the equivalent of agreement among three out of four workers.

Inter-rater agreement was high for the individual assessment items as well as the overall decision. Inter-rater agreement for each assessment item was 90% or higher when examined across all 36 vignettes (see Table 2). Items with lower percent agreement were “inappropriate supervision” and “conditions exist that create substantial likelihood that the child will be harmed due to caregiver neglect.” Situations involving these allegations may require more subjective judgment from screening workers, or workers may have difficulty with these definitions. Discussing these items with screening workers and/or pursuing some targeted case reading may help identify how to increase the consistency of scoring for these items.

It is important to note that a given allegation on the screening assessment was scored for only a subset of the 36 vignettes. For example, physical abuse allegations would have been scored only for reports of physical abuse. Reliability findings were also strong, however, when examined only for cases in which workers scored an item in response to the vignette (for more details, see Appendix A).

<b>Table 2</b>			
<b>Inter-rater Agreement Rates for Pilot Screening and Response Time Assessment Items (Based on 36 Vignettes)</b>			
<b>Assessment Item</b>	<b>Rate of Agreement</b>		
	<b>Average</b>	<b>Minimum</b>	<b>Maximum</b>
<b>Initial Decision</b>	<b>87.9%</b>	<b>53.8%</b>	<b>100.0%</b>
<b>Final Decision</b>	<b>87.6%</b>	<b>50.0%</b>	<b>100.0%</b>
<b>PHYSICAL ABUSE</b>			
Suspicious death of a child due to abuse			
Non-accidental physical injury	94.8%	61.5%	100.0%
Injury inconsistent with explanation	97.7%	66.7%	100.0%
Past injuries that appear suspicious	97.4%	60.0%	100.0%
Giving child toxic chemicals, alcohol, or drugs	96.6%	61.5%	100.0%
Caregiver action that likely caused injury	90.6%	53.8%	100.0%
<b>SEXUAL ABUSE</b>			
Sexual molestation of a child by an adult caregiver, family member, or household member	99.8%	94.4%	100.0%
Sexual act(s) among siblings or other children living in the home that is outside of normal exploration			
Sexual exploitation of a child by an adult caregiver, family member, or household member			
Physical, behavioral, or suspicious indicators consistent with sexual abuse	99.0%	76.9%	100.0%
<b>NEGLECT</b>			
Suspicious death of a child due to neglect			
Signs or diagnosis of non-organic failure to thrive			
Inadequate food/nutrition, or signs of malnutrition	96.0%	52.9%	100.0%
Exposure to unsafe conditions in the home	92.5%	52.9%	100.0%
Inadequate clothing or hygiene	97.1%	52.9%	100.0%
Inappropriate supervision	89.5%	50.0%	100.0%
Failure to protect	97.3%	61.5%	100.0%

<b>Table 2</b>			
<b>Inter-rater Agreement Rates for Pilot Screening and Response Time Assessment Items (Based on 36 Vignettes)</b>			
<b>Assessment Item</b>	<b>Rate of Agreement</b>		
	<b>Average</b>	<b>Minimum</b>	<b>Maximum</b>
<b>Abandonment</b>			
A child of any age has been abandoned			
A child is being discharged from a facility and parents refuse to pick up or plan for the child			
<b>Unattended Child</b>			
A child of any age who is physically, intellectually, or cognitively disabled is left unsupervised or with responsibilities beyond his or her capabilities	98.1%	56.3%	100.0%
A child under the age of 8 has been left alone or in the care of an inappropriate caregiver	96.5%	66.7%	100.0%
A child over the age of 8 has been left alone without support systems for long periods of time or with responsibilities beyond his or her capabilities	97.1%	62.5%	100.0%
<b>Risk of Harm</b>			
Prior death or serious injury of a child due to child abuse or neglect, and a new child is now in the home			
Drug-exposed newborn	98.3%	69.2%	100.0%
Known sexual perpetrator has unsupervised or unrestricted access to child	99.6%	84.6%	100.0%
Conditions exist that create substantial likelihood that the child will be harmed due to caregiver neglect	89.5%	53.8%	100.0%
<b>Medical Neglect</b>			
The unreasonable delay, refusal, or failure on the part of the caregiver to seek, obtain, and/or maintain necessary medical, dental, or mental health care	96.8%	68.8%	100.0%
<b>MENTAL INJURY</b>			
<b>Abuse</b>			
A child has an observable, identifiable, and substantial impairment of his/her mental or psychological ability to function as a result of an act of a parent, permanent or temporary caregiver, or household or family member	98.4%	62.5%	100.0%
<b>Neglect</b>			
A child has an observable, identifiable, and substantial impairment of his/her mental or psychological ability to function as a result of failure to act by a parent, permanent or temporary caregiver, or household or family member	99.0%	81.3%	100.0%

Note: Percentages reported above are based on the number of respondents with the highest level of agreement. Shaded items were not scored for any of the vignettes; therefore, percent agreement could not be calculated.

The kappas derived from this reliability test of the screening assessment indicate that agreement exceeds chance. Fleiss' kappa for the 28 screening assessment items and the decision



was .64, with a confidence interval of .61–.68 (see Table 3). When coding was aggregated into broad categories of maltreatment (physical abuse, sexual abuse, or neglect), the kappa rose to .76 (with a confidence interval of .68–.84). Landis and Koch (1977b) defined thresholds of values equal or greater than .75, between .40 and .75, and below .40 to define excellent, good, and poor agreement beyond chance. Although no empirical basis was provided for the thresholds, these frequently used criteria indicate excellent reliability for the screening and response time assessment.

<b>Table 3</b>	
<b>Reliability Statistics for Screening and Response Time Assessment Items</b>	
	<b>Average Fleiss' Kappa Across 36 Cases (Confidence Interval)</b>
Reliability for the 28 items and decision across screening workers	.64 (.61–.68)
Reliability for maltreatment classifications and decision across screening workers	.76 (.68–.84)

### **C. Summary**

This test of inter-rater reliability among field staff showed that the screening and response time assessment resulted in high levels of agreement. The rate of agreement for the screening decision was 75% or greater for all but four cases.

The level of inter-rater reliability obtained for the screening and response priority assessment is similar to those reported from other studies. For example, in one study involving four raters completing a risk assessment for 80 cases (Baird et al., 1999), three out of four raters reached agreement on 85% of the cases (the average Cohen's kappa, computed for each set of raters, was .562). In a test of a screening assessment used during pediatric exams to identify problematic mother-infant interactions, 15 paired observations of mother and child resulted in a kappa of .74 (Fiese, Poehlmann, Irwin, Gordon, & Curry-Bleggi, 2001).

### **III. EXAMINING PRACTICE WITH PRE- AND POST-IMPLEMENTATION CASE REVIEWS**

The evaluation also included a qualitative review of referrals completed by Maryland workers for a sampled period before and after implementation of the screening and response time assessment. The objective of the case reading was to assess the impact of implementing the screening assessment on the quality of case narratives as well as on the justification for screening and response time decisions. The pre-implementation case review provided a baseline measure of documentation quality. The post-implementation review identified whether workers are completing the screening assessments accurately and whether the quality of documentation had improved since implementation of the assessment.

#### **A. Methodology**

The case review process began with the training of three case readers, which consisted of a brief orientation to the goals of the effort, followed by case readers reviewing the same referrals using the same case reading instrument to refine that instrument and ensure the consistency of case reader scoring. The case reading instrument collected basic information such as the final screening and response time decisions made by the screener and approved by his or her supervisor, as well as information on whether these decisions were supported by the screening narrative and other documentation. In addition, the three CRC staff completed screening and response time assessment for each case read, using the information provided in Maryland's management information system (referred to as MD CHESSIE) screening narrative. The basic questions addressed for each case were the following:

- Were the screening tools completed in a timely manner?
- What was the nature of the allegations recorded by staff?

- To what extent were those allegations supported by the narrative in MD CHESSIE and by the item definitions?
- Were there any allegations that should have been marked but were not?
- What was the screening decision and to what extent was it supported by the MD CHESSIE narrative and the item definitions?
- What was the response time assigned and was it supported by the nature of the referral and policy?

The pre-implementation case reading occurred in September 2008 in non-pilot counties. Reports were randomly selected from a listing of all reports received that month in those counties. The sample included daytime and after-hours calls such as referrals received on weekends and between 4:00 p.m. and 8:00 a.m. on weekdays. Case readers reviewed 164 randomly selected reports received prior to the implementation of the revised screening assessment.

The sample period selected for post-implementation case reading was April 2009. This sample period gave workers additional time after initial training to become familiar with completing the web-based screening assessment. Readers reviewed 244 reports randomly selected from reports with completed screening assessments. Reports received during business and non-business hours were included in both samples.

## **B. Findings**

Table 4 compares findings from the pre-implementation review of reports received to the post-implementation review. Regardless of time period, almost all reports in MD CHESSIE contained narrative supporting the worker's decision whether or not to refer the report for additional in-person investigation. Pre-implementation, 93.3% of reports contained narrative supporting the screening decision. Post-implementation, the corresponding rate was 95.9%.

Reports reviewed post-implementation, however, were significantly more likely to have narrative supporting all allegations indicated and narrative justifying the response time assigned. Pre-implementation, 65.8% of reports contained narrative supporting all allegations indicated in the report and 74.4% contained narrative justification for the response time. Post-implementation, these rates increased to 73.8% and 84.2%. In addition, reports reviewed post-implementation were significantly *less* likely to have narrative support for allegations that were not indicated in the MD CHESSIE record (7.3%, compared to 23.8% of pre-implementation reports reviewed).

<b>Table 4</b>		
<b>Case File Review Findings</b>		
	<b>Pre-implementation Case Reading</b>	<b>Post-implementation Case Reading</b>
<b>Number of Reports Reviewed</b>	<b>164</b>	<b>244</b>
Screening decision supported by narrative	93.3%	95.9%
Allegation types		
a. All supported by narrative*	65.8%	73.8%
b. Some supported by narrative	11.0%	9.0%
c. None supported by narrative	23.2%	17.2%
Allegation not marked that should have been*	23.8%	7.3%
Response time supported by narrative*	74.4%	84.2%

\*Indicates significant difference in rates (z score,  $p < .05$ , two-tailed test).

The case review process identified some implementation issues that affect the consistency of workers' screening decisions. The items discussed with Maryland SSA administrators during debriefings include the following:

1. Implementation fidelity appears to vary by local office. CRC staff read between 8 and 24 cases from each local agency in Maryland. Case reading outcomes indicate that some offices have strong documentation of the allegations, and screening decisions that are consistent with the screening assessment. Other jurisdictions had poor documentation of the referral narrative, and screening decisions that were inconsistent with the screening assessment definitions. The post-implementation case review process included a local briefing on findings in

order to provide each office with feedback to tailor practice improvement efforts to their specific needs. Issues discussed at debriefings provided specific examples of screening decisions that contradicted state policy, and examples of inappropriate narrative and/or decisions. Local administrators who initiated efforts to improve practice after the briefing have noted improvements in both documentation and decision making.

2. Specific types of allegations continue to present a challenge for consistent decision making. After reviewing 400 reports, CRC staff found inconsistent interpretation of many neglect terms. For example, workers consistently screened in reports of child(ren) under 8 whose parent, guardian, or caregiver was not in the immediate vicinity of the child and not supervising the child. Workers were inconsistent, however, in whether they accepted reports of an “unattended child” and “inadequate supervision.” Workers may be interpreting these terms differently, or not referencing the definitions when faced with such a report. CRC staff and SSA administrators are reviewing various possible solutions to this issue, including a possible modification of item definitions.
3. Implementation has identified a small subset of reports for which workers appropriately screened in the report, but only as an override. Two of the over 400 reports read alleged that an adult was encouraging or expecting a child to commit a crime (carry/sell drugs or steal money or valuables). Workers screened these referrals in using the override function (i.e., providing clear documentation of the safety threat). SSA administrators and CRC staff would like to make sure every worker would make the same choice, and are determining whether an assessment definition should be adapted to help ensure this.

## **C. Summary**

The findings from the pre- and post-implementation case readings suggest that implementing the screening and response time assessment improved workers’ documentation of reports and possibly improved the screening decisions they made. In the post-implementation case reading, significantly more reports had narratives that fully justified all types of maltreatment indicated in the report, and had narrative justifying the response time. The review process also identified ways in which administrators can strengthen screening practices.

#### **IV. SOLICITING THE INPUT OF SCREENING STAFF**

As mentioned previously, the objective of this study was to evaluate the impact of having workers complete a structured assessment to help inform their judgment about whether to screen in a reported allegation for in-person investigation. The evaluation of implementation included a web-based survey of screening staff to solicit their feedback about the assessment and its implementation. Soliciting the input of screening workers is crucial to determining whether the assessment was implemented as intended and whether completing the assessment affects how workers take referrals and make screening decisions.

##### **A. Methodology**

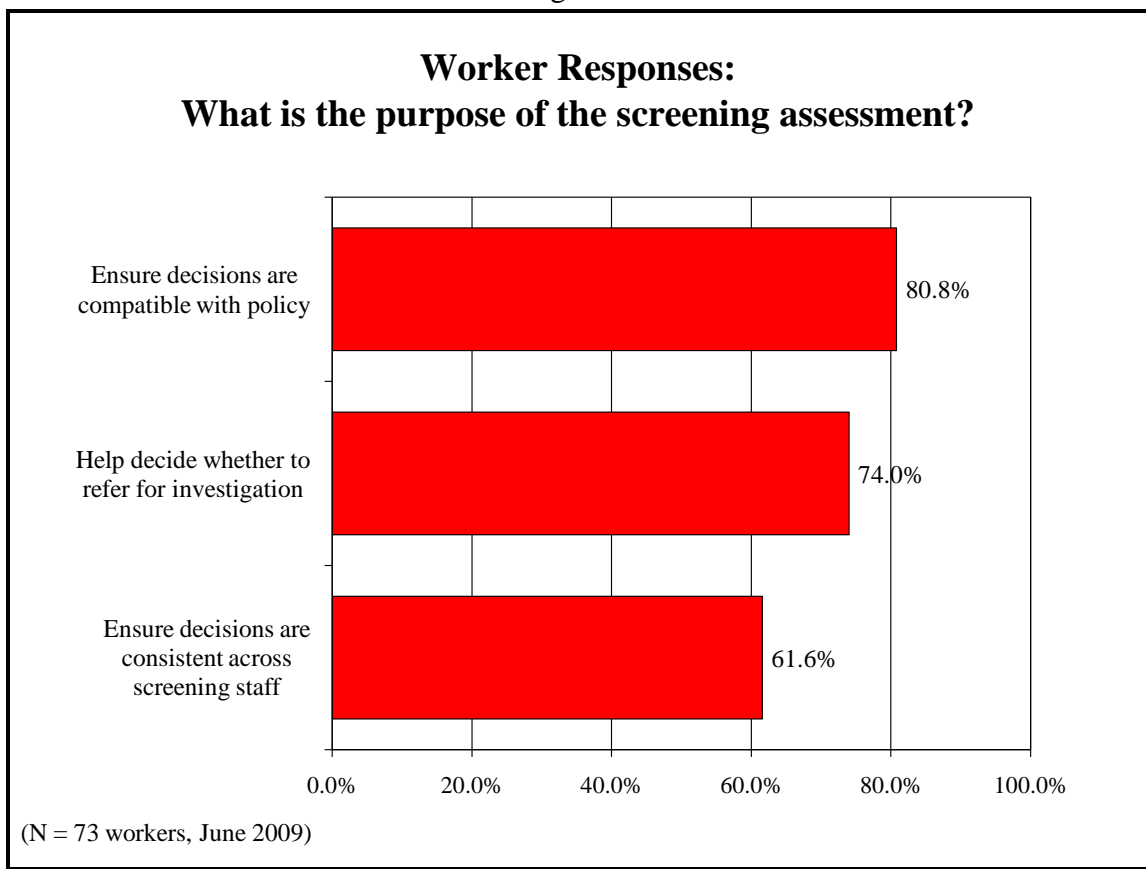
Screening workers were surveyed during the first week of June 2009, and 73 staff from a population of 184 (a 39.7% response rate) responded to provide feedback about the screening assessment. The survey contained structured questions (yes/no and Likert scale) about how staff completed the assessment, how they perceived its usefulness, and how much it affected their decisions to screen in a report (see Appendix B for the full survey). The survey also contained open-ended questions to solicit information without restricting content, and requested basic demographics such as age, experience, and degree.

##### **B. Findings**

Most (90%) of the 73 workers who responded to the web survey were female. Respondents had an average of 7 years' experience in screening (range: .5–30 years) and 13 years of experience in child welfare (range: .5–34 years). The majority (63%) had a master's degree. One fourth (26%) had a bachelor's degree.

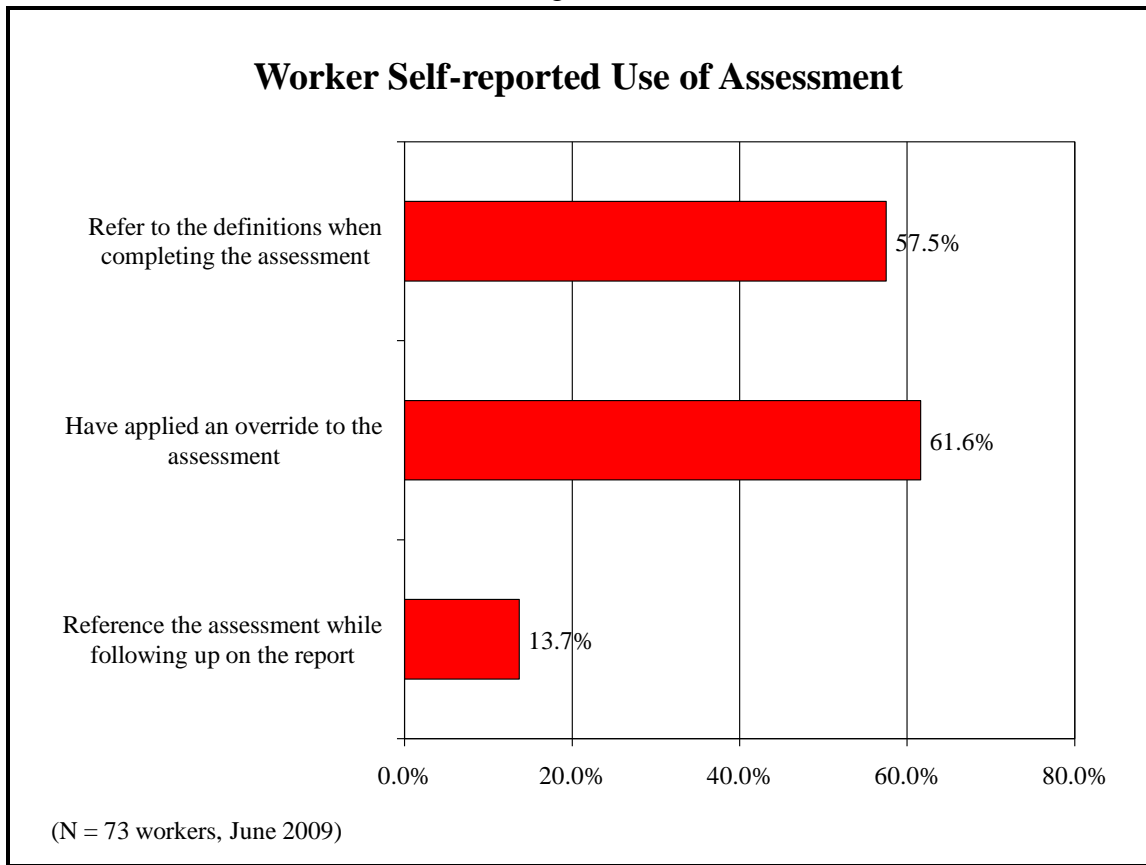
Responding workers were more likely to associate the purpose of the assessment with policy than with decision making. Most (80.8%) knew that the assessment was to translate policy more concisely (see Figure 1). Three fourths of the workers indicated that the assessment was to help decide whether to refer a report for investigation, and just over 60% indicated that it was to increase consistency among decisions.

Figure 1



Nearly two thirds of the responding screening staff indicated that they use the definitions that support the screening assessment and have applied an override to an assessment recommendation (57.5% and 61.6%, respectively; see Figure 2). Thirteen percent of respondents reported having referenced the assessment while following up on a maltreatment report.

Figure 2





On average, responding workers reference the definitions 25% of the time, although some staff reference definitions almost all of the time (see Table 5). Respondents apply overrides infrequently (9% on average). Interestingly, not all workers discuss overrides with a supervisor. On average, responding workers discuss overrides with a supervisor approximately 8% of the time. Policy indicates that workers should review every override with a supervisor.

<b>Table 5</b>		
<b>Self-reported Frequency of Activity</b>		
	<b>Average</b>	<b>Range</b>
Refer to definitions	25%	1–100%
Apply overrides to screening assessment	9%	1–50%
Discuss an override with a supervisor	8%	1–100%

When asked whether they thought the screening assessment helped when making a decision about whether further investigation was needed, approximately 60% of workers who responded to the survey found the assessment helpful. They indicated that it clarified policy in a useful, concise way; that it was a good teaching tool and helped new workers understand policy. They also reported that it was useful in “gray” situations, or when a case was “iffy,”<sup>[31]</sup> because it provides a framework for decisions. One person thought the assessment helped workers “conceptualize why something qualifies for investigation, especially in cases not well defined.”<sup>[45]</sup> Some workers, however, may be completing the assessment to justify a screening decision already made. One worker described the process in this way:

The screening assessment is helpful for supervisors reviewing cases together for approval when there are gray areas for neglect cases and we can find something in the SDM<sup>®</sup> definitions that covers the MD CHESSIE narrative to make a final decision to screen in these types of cases. These types of situations may arise a few times each week.<sup>[63]</sup>

Workers noted that the usefulness of the screening assessment depended on how it was implemented, and that implementation varied by office. One worker noted that “it is helpful,

[but] can be manipulated to meet whatever the screener or screening supervisor wishes the decision to be.” One worker wondered if the agency had a valid and reliable screening assessment.<sup>[27]</sup> Administrators may wish to share report findings with workers as a way of encouraging dialogue.

Practice appears to vary by worker as well as by office. Some workers complete the screening assessment prior to making a decision about whether to forward a report for in-person investigation. Other workers complete the assessment afterward and consider the tool “paperwork.” Workers also varied in the degree to which they referenced the definitions to clarify a situation.

Practice also varied by supervisor. For example, some offices reported that supervisors did not use the screening assessment when reviewing a worker’s decision. Another supervisor, however, reported, “I have asked workers to provide the SDM [screening] assessment when they are in doubt about a case. I have used it to help decide what to do in doubtful cases.”<sup>[59]</sup>

Sometimes workers reported having conflict with and/or not liking the screening assessment, when in fact it was policy that they did not understand or know. For example, one worker reported not liking the assessment because “sometimes we find it conflicts with our social worker thinking. In some abuse reports we would [record] neglect rather than abuse if there were no visible injuries.”<sup>[47]</sup> The complaint lies with neglect policy rather than with the format of the assessment.

A number of workers noted the difficulties of having to work with two pieces of software. One worker said that the assessment was not helpful because it was not in MD CHESSIE: “It has helped a lot in establishing objective definitions of abuse, neglect, etc. But it puts the horse behind the wagon.”<sup>[46]</sup> SSA administrators plan to add the screening and

response time assessment to MD CHESSIE, which will require that screening workers complete it prior to making the screening decision.

### **C. Summary**

More than half (60%) of the 73 workers who responded to the web-based solicitation of feedback found the screening and response time assessment helpful, have referenced the definitions at least once, and have applied an override to the assessment. Workers who found the assessment helpful noted that it clarified policy, provided a decision framework for new workers, and helped when evaluating more difficult reports. Workers who did not consider the assessment helpful said it is additional documentation and that they often complete the assessment after making a decision about the referral. Responding workers noted that the quality of implementation varied across workers and offices.

## **V. DISCUSSION**

### **A. Summary of Findings**

The purpose of evaluating statewide implementation was to determine whether implementing the screening and response time assessment, with training and follow-up case reviews, helped workers reach reliable decisions, and whether implementation improved the narrative support workers provided for screening decisions they made. Evaluation activities included reliability testing of the assessment, a qualitative review of screening decisions conducted before and after implementation, and a survey of workers about the assessment and its implementation. The evaluation sought to answer the following questions:

- Do the screening and response time assessment and its associated definitions help workers obtain consistent screening decisions when provided with the same type and depth of information?

- Has using the screening assessment to structure the decision influenced screening practices?
- Are workers writing more precise narrative?
- Are workers completing the screening and response time assessment as intended according to policy and best practice standards?
- Are workers completing the screening assessment prior to making the decision?

The findings show that completing the screening assessment can be useful to workers (see Table 6). The test of inter-rater reliability among field staff showed that the screening assessment resulted in significant levels of agreement. Participating workers averaged an 88% rate of agreement across 36 vignettes. The rate of agreement for the screening decision was 75% or greater for all but four cases.

<b>Table 6</b>		
<b>Research Question by Evaluative Activity</b>		
<b>Question</b>	<b>Method</b>	<b>Answer</b>
Do the screening and response time assessment and its associated definitions help workers obtain consistent screening decisions when provided with the same information?	Inter-rater reliability testing	Yes
Has using the screening and response time assessment to structure the decision influenced practices involved in screening a report?	Qualitative case review	Yes
Are workers writing more precise narrative?		
Are workers completing the screening and response time assessment as intended according to policy and best practice standards?	Survey of workers	Sometimes
Are they completing it prior to making the screening decision?		

Findings from the pre- and post-implementation case readings suggest that completing the screening assessment improved workers' documentation within the case narrative and possibly improved the screening decisions they made. In the post-implementation case reading,

significantly more reports had narratives that fully justified all types of maltreatment indicated in the report, and had narrative justifying the response time.

Screening staff self-reports confirmed that the screening assessment can aid decision making, but showed that the quality of implementation varied across workers and offices. More than half (60%) of the 73 workers who responded to the web-based solicitation of feedback found the screening assessment helpful and have referenced the definitions at least once. Workers who found the assessment helpful noted that it clarified policy, provided a decision framework for new workers, and helped when evaluating more difficult reports. Workers who did not consider the assessment useful indicated that it is additional documentation and that they often complete the assessment after making a decision about the referral. SSA administrators plan to add the screening and response time assessment to MD CHESSIE, which will require that screening workers complete it prior to making the screening decision.

## **B. Implications**

These findings strongly suggest that the screening assessment and associated definitions, when completed prior to making the screening decision, can help increase the consistency with which screening staff decide whether to refer a maltreatment report for in-person investigation. Workers can increase the consistency of decisions with use of the screening assessment, but this requires consistent implementation and use of definitions.

Research demonstrates that good implementation involves strong administrative support for the practice change, effective training and supervision, and skilled staff (Toth, Manly, & Nilsen, 2008; Mihalic, Irwin, Fagan, Ballard, & Elliott, 2004). SSA administrators may wish to strengthen implementation in some offices by employing efforts used by other jurisdictions, such as the following:

- Emphasize staff use of definitions to promote accurate and consistent assessment scoring. Ensuring that definitions are easily accessible to workers may increase the consistency of decisions.
- Encourage staff and supervisors to reference the screening assessment when reviewing reports. For example, Michigan's Department of Human Services developed a comparative case reading program designed to improve supervisors' evaluation of workers' decision and documentation practices and to improve workers' assessment practices. Supervisors review a sample of case files, and then quality experts review the same file. The supervisors' findings can then be compared to the experts' findings. Findings are reviewed in a summary meeting with supervisors and area managers.
- Solicit questions and concerns from staff to identify areas for follow-up training or additional emphasis. If clarification of policy or intent is needed, managers may want to respond with a written question-and-answer list, ask supervisors to review the subject at a future staff meeting, or revise training materials to include a case example that addresses the issue.
- Continue efforts to improve practice, such as conducting cross-office qualitative case reviews, communicating program successes to staff, and engaging managers and staff in discussions about how to further increase the validity and consistency of decision making.
- Encourage supervisors to routinely review assessment completion and incorporate it into case discussions with workers.
- Ensure that assessment and case action data for CPS reports are easily accessible to staff.

One of the most effective strategies for improving practice identified through the CFSR is monitoring practice with data (Administration on Children and Families, 2006). Regular examination of assessment findings through data reporting can determine how often workers are completing the assessment, what evidence is provided for the screening decision, and if assessment findings are consistent with the final screening decision. If managers had access to such information, they could better monitor and therefore strengthen field practice.

## REFERENCES

- Administration for Children and Families. (2006). *Summary of the results of the 2001–2004 Child and Family Services Reviews*. Washington, D.C.: U.S. Department of Health and Human Services. Retrieved April 30, 2008, from <http://www.acf.hhs.gov/programs/cb/cwmonitoring/results/genfindings04/genfindings04.pdf>.
- Andrews, D.A., Bonta, J., & Wormith, J.S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency*, 52(1), 7–27.
- Baird, C., Wagner, D., Healy, T., & Johnson, K. (1999). Risk assessment in child protective services: Consensus and actuarial model reliability. *Child Welfare*, 78(6), 723–748.
- Campbell, A., Whitehead, J., & Finkelstein, S. (2009). Why good leaders make bad decisions. *Harvard Business Review*, 87(2), 60–66.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
- Fiese, B.H., Poehlmann, J., Irwin, M., Gordon, M., & Curry-Bleggi, E. (2001). A pediatric screening instrument to detect problematic infant-parent interactions: Initial reliability and validity in a sample of high- and low-risk infants. *Infant Mental Health Journal*, 22(4), 463–478.
- Gwets, K.L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *The British Journal of Math and Statistical Psychology*, 61(1), 29–48.
- Hendryx, M., & Rohland, B. (1997). Psychiatric hospitalization decision making by CMHC staff. *Community Mental Health Journal*, 33, 63–73.
- Landis, J.R., & Koch, G.G. (1977a). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 33, 363–374.
- Landis, J.R., & Koch, G.G. (1977b). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Mihalic, S., Irwin, K., Fagan, A., Ballard, D., & Elliott, D. (2004, July). Successful program implementation: Lessons from Blueprints. *The Juvenile Justice Bulletin*.
- Munro, E. (1998). *Understanding social work: An empirical approach*. London: The Athlone Press.
- Rossi, P., Schuerman, J., & Budde, S. (1996). *Understanding child maltreatment decisions and those who make them*. Chicago: Chapin Hall Center for Children.
- Toth, S., Manly, J.T., & Nilsen, W.J. (2008). From research to practice: Lessons learned. *Journal of Applied Developmental Psychology*, 29(4), 317–325.

## **Appendix A**

### **Additional Inter-rater Reliability Findings**



Table A1 shows percent agreement findings limited only to the cases for which an item was scored. Findings for the initial and final screening decision are the same because the decision was noted for every vignette. On average across the 36 vignettes, participating workers reached 87% agreement (column 2) on the screening decision. The lowest percent agreement attained among the workers on any one case was 53%, while the highest attained was 100% agreement (column 4). It is important to note that while the lowest rate of agreement was 53%, the rate was 75% or higher for most of the cases reviewed. Only three cases had lower than 75% agreement for the screening decision, and they were the following (please refer to Table A for a full listing of vignettes):

- Two-parent household (mother and father), newborn. Caller is a social worker at the local hospital. She called last week about a mother who had just given birth. At that time, Mother refused a drug test and did not want her baby tested either. Father became upset and wanted to see a hospital administrator. Today, the social worker is calling to say that she has more information regarding this family. Mother admitted to using Percocet and Xanax throughout the pregnancy. Baby is also showing signs of withdrawal. Baby has not been medicated yet, and it is a borderline case, still waiting on meconium test results. Mother has been compliant with nursing and is remorseful because baby is showing signs of addiction. Mother is bonding with baby. Father reports that they have a nursery set up in the home and his mother is coming to stay with them to help. Both his parents have visited the baby in the hospital several times.
- Single mother, 8- and 5-year-old children (#25). A social worker made a report on one of her open cases. When she visited Girl (8 or 5?) at home today, Girl had a red, puffy, watery eye. When the referent asked her what had happened, Girl stated that her brother was angry and hit her in the eye. When asked to show how he did it, she showed a closed fist. Girl said Brother was mad because she did not flush the toilet. Social worker asked when this happened, and Girl said it was earlier that morning. The social worker asked where Mother was, and Girl said that she was in the kitchen making cereal. When she found out, she yelled at Brother. The social worker also noted that Girl was not wearing a hearing aid she needs because she suffers from hearing loss. Mother said that the battery died last night and she needs to go out and buy a new battery and some spares for next time.

- Single mother, 3- and 4-year-old children (#32). The children’s attorney called in a concern and stated that the home was infested with fleas and mice. The children’s daycare told her that their diaper bag had both mouse droppings and a flea in it. The attorney spoke to Mother, and she admits that they have a mouse problem and that there are a lot of mouse droppings in the apartment. She further stated that they have a flea infestation and that Daughter has red bumps which she thinks are flea bites on her back. Mother has been keeping the bites clean and the pharmacist told her that an over-the-counter itch cream could help. She has been keeping the home clean and has put all food in plastic containers. She vacuums often to try to get rid of the fleas, and washes all the bedding frequently. She has also contacted her landlord and HUD about these issues, but they never return her calls.
- Single mother, 8-year-old child, four other children between infancy and 5 or 6 years old (#9). Neighbor is reporting neglect of Girl (8) and her two minor brothers, names unknown, ages 5 or 6. Reporter states that Girl goes to school “filthy, dirty, and [with] her hair matted” to her head. When asked for more details, Neighbor said that Girl’s clothes have stains on them, and her hair is tangled and snarled. She watches Girl walking to the school bus each day, talking to the other girls, and “her heart bleeds for her.” Reporter states that Girl’s home is also “filthy” in that there are dirty dishes on the kitchen counters and clothing everywhere in the bedrooms. Caller states that she has volunteered to do Girl’s hair but Girl said that she would have to ask her mother for permission because she isn’t allowed in strangers’ houses.

Workers may have assigned different weights to various pieces of information provided in each scenario. It may be that discussing these scenarios during a staff meeting can help identify differences in how workers are assessing situations.

Even when evaluated only for the cases in which an item was scored, findings indicate that the assessment items have high inter-rater reliability. Almost all items averaged 75% or higher rates of agreement. The only exceptions were two neglect items: inadequate food or care, and exposing a child to substances, and workers scored these two items in four or fewer cases.

<b>Table A</b>			
<b>Inter-rater Agreement Rates for Pilot Screening Assessment Items (Based on 36 Vignettes)</b>			
Screening Assessment Item	Rate of Agreement Among Cases With Item Scored Only		
	Average	Number of Cases Assessed	Minimum– Maximum
<b>Initial Decision</b>	<b>87.9%</b>	<b>36</b>	<b>53.8–100%</b>
<b>Final Decision</b>	<b>87.6%</b>	<b>36</b>	<b>50.0–100%</b>
<b>PHYSICAL ABUSE</b>			
Suspicious death of a child due to abuse			
Non-accidental physical injury	73.4%	7	61.5–94.1%
Injury inconsistent with explanation	83.3%	5	66.7–93.8%
Past injuries that appear suspicious	68.4%	3	60.0–76.5%
Giving child toxic chemicals, alcohol, or drugs	59.0%	3	61.5–100%
Caregiver action that is likely to cause injury	77.5%	15	53.8–94.1%
<b>SEXUAL ABUSE</b>			
Sexual molestation of a child by an adult caregiver, family member, or household member	98.1%	3	94.4–100%
Sexual act(s) among siblings or other children living in the home that is outside of normal exploration			
Sexual exploitation of a child by an adult caregiver, family member, or household member			
Physical, behavioral, or suspicious indicators consistent with sexual abuse	81.8%	2	76.9–86.7%
<b>NEGLECT</b>			
Suspicious death of a child due to neglect			
Signs or diagnosis of non-organic failure to thrive			
Inadequate food/nutrition, or signs of malnutrition	64.1%	4	52.9–88.9%
Exposure to unsafe conditions in the home	77.5%	12	52.9–94.1%
Inadequate clothing or hygiene	73.6%	4	52.9–93.8%
Inappropriate supervision	76.3%	16	50.0–94.4%
<b>Failure to Protect</b>	<b>87.7%</b>	<b>8</b>	<b>61.5–93.8%</b>
<b>Abandonment</b>			
A child of any age has been abandoned			
A child is being discharged from a facility and parents refuse to pick up or plan for the child			
<b>Unattended Child</b>			
A child of any age who is physically, intellectually, or cognitively disabled is left unsupervised or with responsibilities beyond his or her capabilities	77.7%	3	56.3–94.4%

<b>Table A</b>			
<b>Inter-rater Agreement Rates for Pilot Screening Assessment Items (Based on 36 Vignettes)</b>			
<b>Screening Assessment Item</b>	<b>Rate of Agreement Among Cases With Item Scored Only</b>		
	<b>Average</b>	<b>Number of Cases Assessed</b>	<b>Minimum– Maximum</b>
A child under the age of 8 has been left alone or in the care of an inappropriate caregiver	81.9%	7	66.7–93.8%
A child over the age of 8 has been left alone without support systems for long periods of time or with responsibilities beyond his or her capabilities	79.1%	5	62.5–92.3%
<b>Risk of Harm</b>			
Prior death or serious injury of a child due to child abuse or neglect, and a new child is now in the home			
Drug-exposed newborn	79.3%	3	69.2–93.8%
Known sexual perpetrator has unsupervised or unrestricted access to child	84.6%	1	--
Conditions exist that create substantial likelihood that the child will be harmed due to caregiver neglect	79.0%	18	53.8–93.8%
<b>Medical Neglect</b>			
The unreasonable delay, refusal, or failure on the part of the caregiver to seek, obtain, and/or maintain necessary medical, dental, or mental health care	85.6%	8	68.8–94.1%
<b>MENTAL INJURY</b>			
<b>Abuse</b>			
A child has an observable, identifiable, and substantial impairment of his/her mental or psychological ability to function as a result of an act of a parent, permanent or temporary caregiver, or household or family member	80.0%	3	62.5–92.3%
<b>Neglect</b>			
A child has an observable, identifiable, and substantial impairment of his/her mental or psychological ability to function as a result of failure to act by a parent, permanent or temporary caregiver, or household or family member	82.3%	2	81.3–83.3%

Note: Percentages reported above are based on the number of respondents with the highest level of agreement. Shaded items were not scored for any of the vignettes; therefore, percent agreement could not be calculated.

## **Appendix B**

### **Web-based Survey of Workers**

SSA recently implemented a Structured Decision Making<sup>®</sup> screening assessment.

The assessment was developed by a workgroup of agency intake workers, supervisors, managers whose task was to translate policy and legislation into a concise assessment to help inform a worker's decision about whether a report indicates credible risk of harm to a child.

We would like to ask you about your experiences with training and use of the screening assessment. All of your responses will remain confidential and if shared will not be attributed to you (anonymity is guaranteed).

Please take a few minutes to give us your opinion.

1. How do you typically use the screening assessment when you complete a report?
2. What is the purpose of the screening assessment? (check all that apply)  
 To ensure screen in decision is compatible with MD policy and procedure  
 To ensure screen in decision is consistent across workers  
 To determine whether a referral requires an investigation  
 Other (If other, please specify: \_\_\_\_\_)
3. In your opinion, does the screening assessment help workers (generally speaking) with decision making and if so, how?
4. How about you specifically? Does the screening assessment help you with decision making and how?
5. Please take a minute to think about the many, varied aspects of your job, and how doing your job might have changed since implementation of the screening assessment and other recent, best practice efforts.

On a scale of 1 to 5 where 1 is much worse and 5 is much better, how has implementing the screening assessment affected the quality of your work?

	Much Worse	Somewhat Worse	Neither Better nor Worse	Somewhat Better	Much Better
a. How and what you collect from a reporter	1	2	3	4	5
b. How you narrate your findings on a report	1	2	3	4	5
c. How you decide to screen in a report	1	2	3	4	5

6. Does the screening assessment make your job harder or easier?  
 Makes my job easier  
 Neither harder nor easier  
 Makes my job harder

The next few questions are about how you physically use the screening assessment.

7a. Do you complete the assessment on paper, prior to doing it on the computer?

1. Yes
2. No

7b. If yes, approximately how long in minutes does it take you to note responses on the paper assessment?

\_\_ minutes

8a. Do you complete the assessment by computer?

1. Yes
2. No

8b. If yes, approximately how long in minutes does it take you to complete the computer-based assessment?

\_\_ minutes

9a. Do you refer to the definitions when completing the screening assessment?

1. Yes
2. No

9b. If yes, for what percentage of assessments do you refer to the definitions?

\_\_ %

10. Do you ever reference the screening assessment as you follow up on reports of maltreatment? (For example, does it affect how you follow up on a report, whether you call collaterals and who you call?)

1. Yes
2. No

11a. Have you ever applied an override to a screening assessment?

1. Yes
2. No

12. If you have ever applied an override:

For what percentage of reports do you apply overrides?

\_\_ %

What percentage of overrides did you discuss with your supervisor?

\_\_ %

13. Are there other ways in which the structured screening assessment assisted you in your work? How so?

We would like you to answer a few more questions to help make sure we received responses from a variety of workers:

14. How many years have you worked in Intake? \_\_\_\_

15. How many years have you worked in a child welfare agency in direct service? \_\_\_\_

16. How many years have you worked in another public service, such as juvenile justice, adult corrections, mental health services, or adult protective services? \_\_\_\_

17. Please mark the highest degree you have earned:

BA/BS

MSW

MPH

PhD

Other (If other, please specify: \_\_\_\_\_)

18. What is your age in years? \_\_\_\_

19. Please indicate your gender:      1. Male      2. Female

Thank you for completing this survey. Your answers will help us better understand the implementation of intake assessments in Maryland.



## **Appendix C**

### **SDM<sup>®</sup> Child Abuse and Neglect Screening and Response Time Assessment Tool and Definitions**

MARYLAND SOCIAL SERVICES ADMINISTRATION  
STRUCTURED DECISION MAKING®  
CHILD ABUSE AND NEGLECT SCREENING AND RESPONSE TIME ASSESSMENT

r: 10-08

---

**Referral Name:** \_\_\_\_\_ **Referral Date:** \_\_\_\_/\_\_\_\_/\_\_\_\_  
**CHESSIE Referral ID:** \_\_\_\_\_ **Referral Time:** \_\_\_\_:\_\_\_\_ a.m./p.m.  
**Jurisdiction:** \_\_\_\_\_

---

**SECTION 1. MALTREATMENT TYPE**

---

**PHYSICAL ABUSE**

- Suspicious death of a child due to abuse.
- Non-accidental physical injury.
- Injury inconsistent with explanation.
- Injury that appears suspicious.
- Giving child toxic chemicals, alcohol, or drugs.
- Caregiver action that likely caused injury.

---

**SEXUAL ABUSE**

- Sexual molestation of a child by an adult caregiver, family member, or household member.
- Sexual act(s) among siblings or other children living in the home that is outside of normal exploration.
- Sexual exploitation of a child by an adult caregiver, family member, or household member.
- Physical, behavioral, or suspicious indicators consistent with sexual abuse.

---

**NEGLECT**

**General Neglect**

- Suspicious death of a child due to neglect.
- Signs or diagnosis of non-organic failure to thrive.
- Inadequate food/nutrition, or signs of malnutrition.
- Exposure to unsafe conditions in the home.
- Inadequate clothing or hygiene.
- Inadequate supervision.
- A child is being discharged from a facility and parents refuse to participate in planning for the child.

**Failure to Protect**

- The caregiver does not intervene despite knowledge (or reasonable expectation that the caregiver should have knowledge) that the child is being harmed (includes physical or sexual abuse, neglect, or mental injury) by another person.

**Abandonment**

- A child of any age has been abandoned.

**Unattended Child**

- A child of any age who is physically, intellectually, or cognitively disabled is left unsupervised or with responsibilities beyond his or her capabilities.
- A child under the age of 8 has been left alone or in the care of an inappropriate caregiver.
- A child over the age of 8 has been left alone without support systems for long periods of time or with responsibilities beyond his or her capabilities.

**Risk of Harm**

- Prior death or serious injury of a child due to child abuse or neglect, and a new child is now in the home.
- Drug-exposed newborn.
- Known sexual perpetrator has unsupervised or unrestricted access to child.
- Child's basic needs are likely to be unmet due to caregiver impairment.

**Medical Neglect**

- The unreasonable delay, refusal, or failure on the part of the caregiver to seek, obtain, and/or maintain necessary medical, dental, or mental health care.

---

Structured Decision Making® and SDM®  
Registered in the U.S. Patent and Trademark Office

---

**MENTAL INJURY****Abuse**

- A child has an observable, identifiable, and substantial impairment of his/her mental or psychological ability to function as a result of an act of a parent, permanent or temporary caregiver, or household or family member.

**Neglect**

- A child has an observable, identifiable, and substantial impairment of his/her mental or psychological ability to function as a result of failure to act by a parent, permanent or temporary caregiver, or household or family member.

---

**SECTION 2. RECOMMENDATION AND OVERRIDES**

---

**Initial Screening Recommendation**

- Screen out (no maltreatment type is marked)  
 Screen in (one or more maltreatment types are marked)

**Overrides**

- No overrides apply.
- Screen out: One or more maltreatment types are marked, but referral will be screened out (*mark all that apply*).
- Insufficient information to locate child/family.
  - Information forwarded to another jurisdiction.
  - Historical information, victim is now an adult, and no children are in the care of alleged perpetrator.
  - Other (specify): \_\_\_\_\_
- Screen in: No maltreatment type is marked, but referral will be opened and assigned for child protective services (CPS) investigation (*mark all that apply*).
- Court order for an investigation.
  - Other (specify): \_\_\_\_\_

**Final Screening Decision** (*after consideration of overrides*)

- Screen out: No maltreatment type is marked AND no screen-in overrides apply.
- Screen in: At least one maltreatment type or screen-in override is marked AND no screen-out overrides are marked. Complete Section 3, Response Time Decision.

---

**SECTION 3. RESPONSE TIME DECISION** (*complete for all screened in reports*)

---

- Immediate response required based on one or more criteria below** (*mark all that apply*):
- Child fatality or near fatality where abuse/neglect is suspected.
  - Serious injury to child and other children remain in home.
  - Child left alone/abandoned and requires immediate care.  
Age of youngest child in years: \_\_\_\_\_
  - Allegation of child abuse or neglect in an out-of-home setting.
  - Other (specify): \_\_\_\_\_
- No immediate response criteria exist and allegations include the following:**
- Physical abuse—**response within 24 hours**
  - Sexual abuse—**response within 24 hours**
  - Neglect—**response within 5 days**
  - Mental injury—**response within 5 days**
  - Screen-in Override**

---

**SECTION 4. DATA COLLECTION**

---

- Yes  No Report includes allegation of abuse in which child is under 1 year old and has received corporal punishment.

Notes: \_\_\_\_\_

Worker: \_\_\_\_\_ Date: \_\_\_\_/\_\_\_\_/\_\_\_\_

Supervisor: \_\_\_\_\_ Date: \_\_\_\_/\_\_\_\_/\_\_\_\_

**MARYLAND SOCIAL SERVICES ADMINISTRATION  
STRUCTURED DECISION MAKING®  
CHILD ABUSE AND NEGLECT SCREENING AND RESPONSE TIME ASSESSMENT  
DEFINITIONS**

**SECTION 1. MALTREATMENT TYPE**

**PHYSICAL ABUSE** is an act of commission by a parent, caregiver, or other household member in which a child sustained, or is likely to have sustained, a physical injury. Injury to the child may be current or may have occurred in the past. Injuries need not be visible for the report to be accepted for investigation.

Suspicious death of a child due to abuse. Caregiver, family member, or other household member allegedly abused and caused the death of a child.

Non-accidental physical injury. A non-accidental physical injury to a child is one that is inflicted by the caregiver, regardless of motive. Include injuries that result from a domestic violence incident, but exclude injuries that result from sexual acts (recorded under sexual abuse). Injuries include but are not limited to the following:

- Bruises and lacerations;
- Burns and/or scalds (reddening or blistering of the tissue through application of heat by fire, chemical substances, cigarettes, matches, electricity, scalding water, friction, etc.);
- Injuries to bone, muscle, cartilage, or ligaments (fractures, dislocations, sprains, strains, displacements, hematomas, etc.);
- Head injuries; and
- Internal injuries.

Injury inconsistent with explanation. Child has a concerning physical injury and caregiver or child provides details of an incident that are inconsistent with the injury.

Injury that appears suspicious. These may include head injuries, bruises, lacerations, internal injuries, burns, scalds, fractures, dislocations, sprains, strains, displacements, and hematomas. Injury or injuries may be current or may be in different stages of healing.

Giving child toxic chemicals, alcohol, or drugs that caused or could cause harm, such as poison, gasoline, kerosene, bleach, cleaning agents, or an inappropriate dosage of medication.

Caregiver action that likely caused injury. It is not necessary for a reporter to determine that an injury occurred. Examples of caregiver action that likely caused injury include but are not limited to the following:

- Shaking, shoving, or throwing an infant or young child.

- Inappropriate physical discipline, such as the use of choking, torture, suffocation, tying child up, or the use of dangerous objects (e.g., whips) to strike child.
- Dangerous behavior toward the child or in immediate proximity of the child. Consider combination of child location, type of incident (e.g., pushing, throwing objects, use of firearm), and child vulnerability.
- The caregiver has made threats to cause physical harm to the child that, if carried out, would constitute child abuse; and it is likely that without intervention the caregiver will carry out these threats.
- Munchhausen Syndrome by Proxy: Caregiver has harmed a child in order to secure for him/herself the role of life-saver and/or life protector. Caregiver may have created or exaggerated the child's symptoms in several ways, including lying about symptoms; altering diagnostic tests (such as contaminating a urine sample); falsifying medical records; or inducing symptoms through various means, such as poisoning, suffocating, starving, or causing infection.

**SEXUAL ABUSE** includes sexual molestation or exploitation, whether physical injuries are sustained or not, by a parent, guardian, household or family member, or someone who has temporary or permanent care and custody or responsibility for supervision of a child.

Sexual molestation of a child by an adult caregiver, family member, or household member based on verbal or nonverbal disclosure, medical evidence, or credible witnessed account. Examples may include the following:

- Perpetrator touching a child's sexual areas (e.g., genitals, buttock, breast) with any part of his/her body or an object.
- Perpetrator touching a child with his/her sexual areas or having a child touch his/her sexual areas.
- Perpetrator exposing his/her sexual areas to a child.
- Perpetrator observing or trying to observe a child's sexual areas.

Sexual act(s) among siblings or other children living in the home that is outside of normal exploration. Children living in the home engage in sexual behavior that is outside of normal exploration and developmentally appropriate behavior. This includes sexual acts among children that involve coercion, threats, or violence.

When episodes of sexual activities between or among children in a family or household are reported, and where all parties to the incident or incidents are under 18, consider the following factors in making a determination of whether the referral should be categorized for investigation as sexual abuse:

- The existence of a size, age, developmental, and/or power differential between the alleged perpetrator and the alleged victim that favors the perpetrator.
- The status of the parties: for example, if one of the individuals is in the caregiver's role on a regular or occasional basis or when the alleged abuse occurred.
- The psychological condition of each child; for example, has either child been diagnosed with a mental illness?
- Did the incident(s) seem primarily related to lack of appropriate supervision or child management, rather than possible sexual abuse?

Sexual exploitation of a child by an adult caregiver, family member, or household member. The caregiver, family member, or household member involves the child in obscene acts; engages the child in prostitution; or allows, permits, encourages, or engages in obscene or pornographic display, photographing, filming, or depiction of a child as prohibited by law. A caregiver, relative, or household member has made sexual advances toward a child including but not limited to asking the child to perform sexual acts or describing sexual activities.

Physical, behavioral, or suspicious indicators consistent with sexual abuse. Suspicious indicators include but are not limited to the following:

- A pre-adolescent child has a sexually transmitted infection or symptoms, or genital or anal area is red, torn, or chafed.
- A pre-adolescent child has initiated sexual acts or activities with caregivers, family members, or a peer that are outside age-appropriate exploration or development, and this has led to a concern that he or she is a victim of other sexual abuse.
- A child is complaining of pain in the genital or anal area AND there are other indications of sexual abuse.

**NEGLECT** is an act of omission by a parent, guardian, caregiver, or legal custodian in failing to provide for the adequate care and attention of the child's needs, resulting in physical or mental harm to the child or substantial risk of physical or mental harm to the child.

### **General Neglect**

Consider age/developmental status of the child. Injury need not have occurred.

Suspicious death of a child due to neglect. There is concern that caregiver's neglect led to the death of a child.

Signs or diagnosis of non-organic failure to thrive. The child has been diagnosed as having non-organic failure to thrive or has indicators of failure to thrive, and/or a caregiver's inattention is causing the condition to worsen.

Inadequate food/nutrition, or signs of malnutrition. The caregiver does not provide sufficient food to meet minimal nutritional requirements for the child. The child experiences significant lack of food, or unmitigated hunger due to lack of food.

Exposure to unsafe conditions in the home. The child's house is significantly unsanitary and/or contains hazards that have led or could lead to injury or illness of the child if not resolved. Examples may include housing that is an acute fire hazard or has been condemned, methamphetamine production in the home, exposed heaters, gas fumes, faulty electrical wiring, no utilities (e.g., heat, water, electricity), broken windows or stairs, vermin, human or animal excrement, unguarded weapons, and accessible drugs or hazardous chemicals.

Inadequate clothing or hygiene. Caregiver has failed to meet a child's basic needs for clothing and/or hygiene to the extent that the child's daily activities are adversely impacted and/or the child develops or suffers worsening of a medical condition (e.g., sores, infection, severe diaper rash, physical illness, serious harm, hypothermia, or frostbite).

Inadequate supervision. Child is not supervised to the extent that he/she has been injured, or avoided injury despite lack of attention or supervision by the caregiver.

A child is being discharged from a facility and parents refuse to participate in planning for the child. Parent refuses to participate in planning for child AND refuses to accept the child back into his/her home AND refuses to allow for voluntary placement of the child. Consideration should be given to the need for continued treatment; risk to the child; family, community, and resource availability.

## **Failure to Protect**

The caregiver does not intervene despite knowledge (or reasonable expectation that the caregiver should have knowledge) that the child is being harmed (includes physical or sexual abuse, neglect, or mental injury) by another person. If the person causing harm is a caregiver, parent, or other household member, consider screening in a referral of physical abuse on the alleged maltreater.

## **Abandonment**

A child of any age has been abandoned. Unharmed infants (10 days or younger) surrendered under Maryland's Safe Haven statute should not be investigated unless there are questions regarding the infant's care since birth. The law allows for unharmed infants 10 days old or younger to be relinquished to a responsible individual who in turn must give the infant to law enforcement or a hospital. The infant may also be relinquished directly to law enforcement or a hospital by the parents (mother, father, or both) of the child. Such situations should not be

handled as investigations and should be referred to a non-child protective services (CPS) intervention.

### **Unattended Child**

A child of any age who is physically, intellectually, or cognitively disabled is left unsupervised or with responsibilities beyond his or her capabilities.

A child under the age of 8 has been left alone or in the care of an inappropriate caregiver. A child under the age of 8 has been left alone or left in the care of a person who is not reliable or is under the age of 13.

A child over the age of 8 has been left alone without support systems for long periods of time or with responsibilities beyond his or her capabilities.

- A child aged 8 through 12 is left alone in the following circumstances:
  - » For longer than brief periods, without support systems that should include phone numbers of parents, other family members, or neighbors; information about personal safety; and what to do in an emergency;
  - » To care for children under the age of 8; or
  - » With responsibilities beyond his or her capabilities.
- A child age 12 or over is left alone in the following circumstances:
  - » For long hours, including overnight, without support systems that should include phone numbers of parents, other family members, or neighbors; information about personal safety; and what to do in an emergency; or
  - » With responsibilities beyond his or her capabilities.

### **Risk of Harm**

Conditions exist that create a substantial likelihood that the child will be harmed due to caregiver's neglect. It is not necessary for injury to have occurred.

Prior death or serious injury of a child due to child abuse or neglect, and a new child is now in the home. There is credible information that a current caregiver was responsible for the death or serious injury of a child due to neglect/abuse AND there is now a new child living in the home.

Drug-exposed newborn. Infant is born drug-exposed, as indicated by a positive toxicology screen for illegal drugs, symptoms of withdrawal, or other indicators as determined by medical personnel; AND mother's behaviors or condition indicate newborn may be at risk of abuse or neglect. These behaviors or indicators may include but are not limited to previous delivery of drug-exposed newborn; lack of consistent prenatal care; prior CPS history; no preparations for the child's arrival; and lack of responsiveness to the child's needs while in the hospital/healthcare setting. See circular letter SSA# 08-6 for further details.



Known sexual perpetrator has unsupervised or unrestricted access to child. Caregiver knowingly allows a person with history of sexual offenses against minors to have unsupervised and/or unrestricted access to child.

Child's basic needs are likely to go unmet due to caregiver impairment. Caregiver's ability to parent appears to be substantially impaired to the extent that the caregiver would be unable to respond to or meet the basic needs of the child (food, clothing, shelter, education, healthcare) and the caregiver has not made other arrangements for supervision or care of the child. Impairment may be caused by mental or physical health conditions or active substance abuse.

## **Medical Neglect**

The unreasonable delay, refusal, or failure on the part of the caregiver to seek, obtain, and/or maintain necessary medical, dental, or mental health care when caregiver knows, or should reasonably be expected to know, that such actions may cause adverse impact on the child's health and welfare. Such actions may include but are not limited to the following:

- Missed appointments, therapies, or other necessary medical and/or mental health treatments;
- Withholding or failing to obtain or maintain medically necessary treatment for a child with life-threatening, acute, or chronic medical or mental health conditions; and
- Failing to provide comfort measures to infants and children with life-ending conditions.

**MENTAL INJURY** is an observable, identifiable, and substantial impairment of a child's mental or psychological ability to function as a result of child abuse or neglect.

## **Abuse**

A child has an observable, identifiable, and substantial impairment of his/her mental or psychological ability to function as a result of an act of a parent, permanent or temporary caregiver, or household or family member. These acts may include the following:

- Implied or overt threats of death or serious injury of the child or others;
- Implied or overt threats in the form of pet or animal torture; or
- Constant denigration.

## **Neglect**

A child has an observable, identifiable, and substantial impairment of his/her mental or psychological ability to function as a result of a failure to act by a parent, permanent or temporary caregiver, or household or family member. This may include extensive emotional or physical isolation or confinement, or severe lack of engagement or stimulation.

## SECTION 2. RECOMMENDATION AND OVERRIDES

### Initial Screening Recommendation

Screen out: No maltreatment type is marked.

Mark this decision if no maltreatment type in Section 1 is marked, which means that the referral does not meet statutory requirements for an in-person response.

Screen in: One or more maltreatment types are marked.

Mark this decision if any maltreatment type in Section 1 is marked, which means that at least one reported allegation meets statutory requirements for an in-person response.

### Overrides

No overrides apply.

Screen out: One or more maltreatment types are marked, but referral will be screened out.  
Indicate the reason.

- Insufficient information to locate child/family. The caller was unable to provide enough information about the child's identity and/or location to enable an in-person response. Do not mark this item if partial information is available. Screener should either follow up on information to establish child's identity/location or forward screened-in referral for investigation.
- Information forwarded to another jurisdiction. Abuse or neglect occurred in another county or state. Information forwarded to appropriate jurisdiction for response.
- Historical information, victim is now an adult, and no children are in the care of an alleged perpetrator. Identified victim is now an adult, and alleged perpetrator is not a caregiver, household member, or family member of minor children.
- Other (specify).

Screen in: No maltreatment type is marked, but referral will be opened and assigned for child protective services (CPS) investigation.

Mark this decision if no maltreatment types in Section 1 are marked, which means that the referral does not meet statutory requirements for an in-person response. However, a referral will be opened and assigned for investigation for one or more of the following reasons:

- Court order for an investigation.
- Other, specify.

## **Final Screening Decision** (after consideration of overrides)

Screen out: No maltreatment type is marked AND no screen-in overrides apply.

Mark this decision if no maltreatment type in Section 1 is marked, which means that the referral does not meet statutory requirements for an in-person response, AND no screen-in overrides in Section 3 are marked.

Screen in: At least one maltreatment type or screen-in override is marked AND no screen-out overrides are marked.

Mark this decision if any criteria in Section 1 are marked, which means that at least one reported allegation meets statutory requirements for an in-person response, or at least one screen-in criteria was identified AND no screen-out criteria were marked. For all referrals in which the final screening decision is to screen in, a response time must be identified.

## **SECTION 3. RESPONSE TIME DECISION**

For all screened-in referrals, review criteria for immediate response and mark all that apply. If any apply, immediate response is required by the local agency. If no immediate response criteria exist, mark the type(s) of maltreatment that were identified in the allegations. Response time will be based on the most severe type of maltreatment alleged.

## **SECTION 4. DATA COLLECTION**

For all screened-in referrals, each question must be answered either yes or no.

Report includes allegation of abuse in which child is under 1 year old and has received corporal punishment. Mark yes if the report met criteria for abuse or neglect and involved a child under 1 year old receiving corporal punishment.

## **Appendix D**

### **Findings From the Field Test**

# **A Field Test of the Child Abuse and Neglect Screening and Response Time Assessment**

A Report Produced for the Maryland Social Services Administration

November 2008

Kristen Johnson  
Deirdre O'Connor



Children's Research Center, a private nonprofit social research organization and  
division of the National Council on Crime and Delinquency  
426 South Yellowstone Drive, Suite 250  
Madison, WI 53719  
(608) 831-1180  
[www.nccd-crc.org](http://www.nccd-crc.org)

## Table of Contents

INTRODUCTION .....	D3
DESCRIPTION OF THE FIELD TEST.....	D4
PRE- AND POST-IMPLEMENTATION CASE READING .....	D5
Methods.....	D6
Findings.....	D7
Summary of Findings.....	D10
INTER-RATER RELIABILITY TESTING.....	D11
Methods.....	D11
Findings.....	D12
Summary.....	D15
SUMMARY OF FINDINGS FROM THE FIELD TEST .....	D16
Review of Changes to Select Screening Items .....	D16
Review of Statewide Implementation Plan.....	D18
REFERENCES .....	D19

## INTRODUCTION

The goal of implementing the Structured Decision Making<sup>®</sup> (SDM) screening assessment in Maryland is to increase consistency in screening decisions between workers and between local agencies. These screening decisions also need to be consistent with Code of Maryland regulations (COMAR) and with Social Services Administration (SSA) state policy. The implication of unreliable and/or inaccurate case assessments is that families may receive very different treatment recommendations based on the worker assigned to their case, therefore making it difficult to achieve equity and fairness.

The decisions a social worker makes (e.g., whether services should be offered, or whether treatment must be inpatient) influence the course of an individual's life. Unfortunately, research shows that case decisions based on clinical judgment have little predictive validity (Andrews, Bonta, & Wormith, 2006) and are unreliable among the most qualified and experienced workers. For example, one study to test the inter-rater reliability of child protective services (CPS) workers' decisions regarding family services and child removals from home compared the case recommendations of child welfare workers and those of nationally recognized practitioners (Rossi, Schuerman, Budde, & 1996). Both groups had low percent agreement regarding the decision whether or not to place a child, and percent agreement declined when in-home services was added as an option.

A similar study showed that, while workers demonstrated low reliability for case actions, the same workers had full confidence in the rightness of their decisions. Hendryx and Rohland (1997) compared the inter-rater reliability among 82 staff from a community-based mental health center using two cases and found that staff varied widely in their treatment recommendations, though all were very confident in the accuracy of their decision making. Other research demonstrated similar findings (Munro, 1998).

Administrators of Maryland's SSA recognize the importance of reliable decision making in social services, and are therefore determining how best to ensure the consistency of worker decisions about whether or not to screen in a report. They contracted with the Children's Research Center (CRC), a nonprofit social research organization and division of the National Council on Crime and Delinquency, to develop a structured screening assessment that included a comprehensive field test to evaluate whether the assessment was meeting its goals. Implementation of the final screening assessment will also be evaluated on a statewide basis to ensure that the assessment is being used correctly and that workers are being supported in their decisions.

This report summarizes findings from the 60-day field test of the SDM<sup>®</sup> screening assessment. Evaluation focused on how well the screening assessment helped workers reach reliable decisions and could provide narrative support for those decisions. Evaluation of whether these goals were met included case reading conducted before and after field test implementation, and reliability testing among a sample of workers. These activities and the findings are described in detail below.

## **DESCRIPTION OF THE FIELD TEST**

Three local Department of Social Services (DSS) agencies volunteered to participate in a field test of the screening assessment:

- Baltimore City, which comprises approximately 60% of Maryland's CPS workload;
- Anne Arundel County, which includes the state capital of Annapolis as well as rural and suburban areas; and
- Montgomery County, which is a large urban and suburban county immediately adjacent to the District of Columbia.



In Baltimore City and Anne Arundel County, all screeners participated in the field test and completed the SDM screening assessment on all referrals of CPS allegations during the 60-day field test. Montgomery County chose to train three screeners, but, due to illness and other absences, had limited opportunity to utilize the assessment during the field test.

CPS screeners from Baltimore City and Montgomery County were trained on the use of the SDM screening assessment on June 26–27 and July 1, 2008, and began using a paper version of the assessment immediately after training. A web-based version of the SDM screening assessment was made available on July 28, which allowed for consistent data collection from all field test sites. Baltimore City and Montgomery County screeners started using the web-based screening assessment on July 28; screeners in Anne Arundel County were trained on July 29, and started using the web-based screening tool immediately.

## **PRE- AND POST-IMPLEMENTATION CASE READING**

The objective of the case reading was to assess the impact of implementing the screening assessment on the quality of case narratives as well as on the justification for screening and response time decisions. Only Baltimore City intakes were reviewed. The pre-implementation case review provided a baseline measure of documentation quality. The post-implementation review identified whether the SDM screening assessments were being completed correctly and whether the quality of documentation had improved since implementation of the SDM screening assessment.

## Methods

The case reviews were conducted by two experienced CRC staff during the fall of 2008. These staff initially read cases together and discussed them to ensure a reliable and valid analysis of intake decisions.

CRC staff also developed a case reading instrument that helped ensure a structured and systematic collection and review of case information. The case reading instrument collected basic information such as the final screening and response time decisions made by the screener and approved by his or her supervisor, as well as information on whether these decisions were supported by the intake narrative and other documentation. If the decisions were not supported by the narrative, the case reader was expected to provide an explanation. The instrument also included questions on how the case reader would have completed the SDM screening assessment and if this completion led to the same screening and response time decisions made by the Baltimore City screener and supervisor. When used for the post-implementation case reading, there were specific questions about whether the case reader completed the SDM screening assessment identically to the screener; if not, specific details were documented as to what was different. All case records were reviewed in the Maryland SACWIS system, MD CHESSIE.

The sample period selected for the pre-implementation case reading was May 2008, which preceded all discussions between Baltimore City supervisors and managers of the screening units and CRC staff on the intent and structure of the SDM screening assessment. Specific dates were selected to ensure that adequate daytime and after-hours cases were included in the sample. After-hours calls included referrals that came in on weekends and between 4:00 p.m. and 8:00 a.m. on weekdays. Within each selected date, every third case record was included, regardless of screening decision. A total of 50 case records were reviewed.

The sample period selected for post-implementation case reading was September 2008. This sample period allowed for adequate time after initial training for workers to become familiar with completing the web-based SDM screening assessment. Again, specific dates were selected to ensure that adequate daytime and after-hours cases were included in the sample. Within each selected date, every third case record was included, regardless of screening decision. A total of 60 case records were reviewed.

<b>Table 1</b>		
<b>Sample Details for the Baltimore City Field Test Case File Reviews</b>		
<b>Review Period</b>	<b>Dates of Case Reading</b>	<b>Total Number of Cases Read</b>
Pre-implementation	May 2008	50
Post-implementation	September 2008	60

## **Findings**

While representative of work completed by the screening units in Baltimore City, the samples for both pre- and post-implementation case reading were not large enough to provide statistically significant quantitative analysis. Changes over time will be discussed through the use of percentages and raw numbers; however, these findings should be considered qualitative in nature.

From the pre-implementation case review, 17 of the 50 referrals were screened out; in four of these, case records the narrative did not support the screen-out decision. The case readers identified reasons why the agency should have responded to the referral; however, in two of these cases, the agency did respond, though not with an investigation. In one case, an agency worker responded to the home and then determined that no credible information existed to support an allegation; the other case documented the need for a courtesy visit, with explicit instructions to refer back to CPS screening if evidence of physical abuse existed. Both of these

cases documented a response by the agency, in one case an immediate response to see if the children were safe.

In two of the four screened-out cases which should have been screened in, narrative appeared to meet the definition for unattended child (in one, three children under 8 years of age were playing in the street, unattended by an adult; in the other, a grandmother died, leaving two teens without supervision). The other two cases appeared to meet the “physical abuse, risk of harm” definition on the SDM screening tool. Subsequent to the case reading, the “physical abuse risk of harm” item was changed and it appeared that neither of these cases met the new definition.

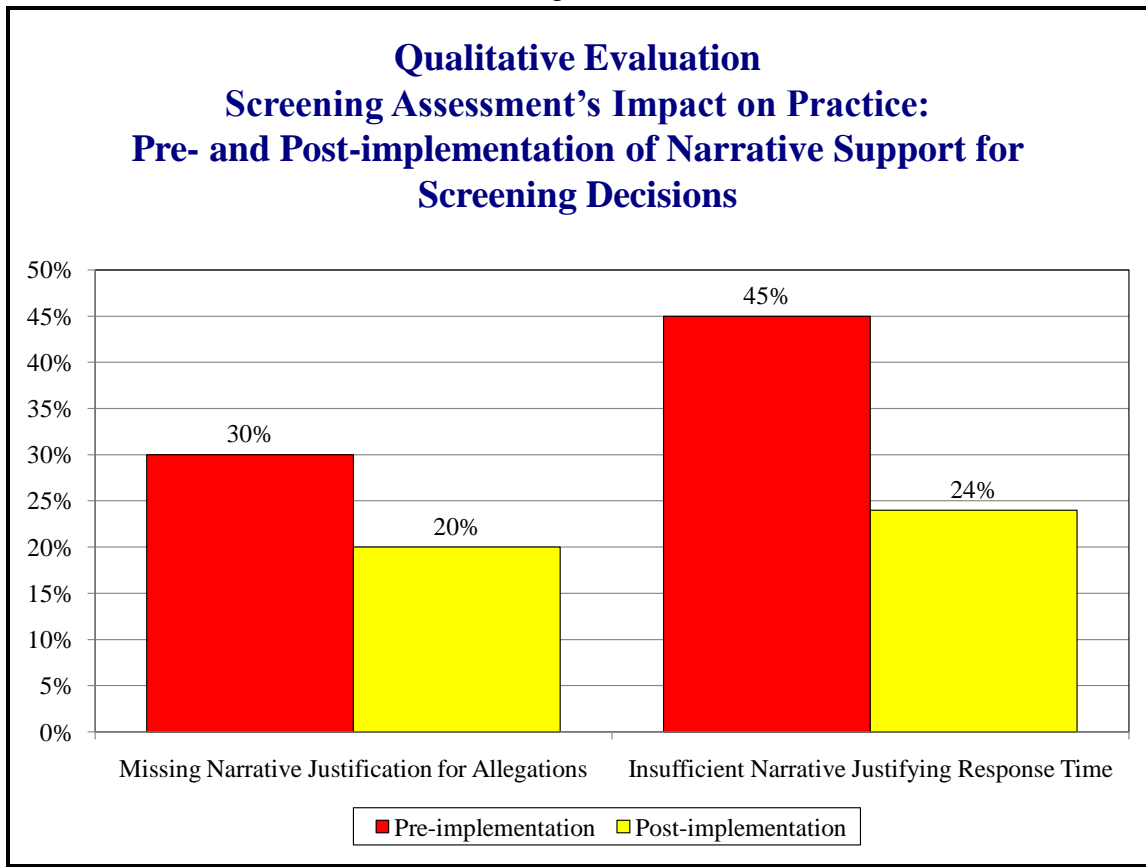
Of the 33 case records reviewed that resulted in a screen-in decision, five cases (15%) appeared to have no allegations supported by case narrative and five cases (15%) had only some of the allegations supported by case narrative. Types of allegations not supported by narrative were primarily neglect allegations, including lack of supervision, failure to protect, and “risk of harm, drug-exposed newborn.”

Of the 33 case records reviewed that resulted in a screen-in decision, case readers identified problems with the response time assigned in 15 cases (45%). Five of these cases were assigned an immediate or priority response when the need for such a response was not clearly identified in the case narrative. In only two of these cases, the narrative indicated a need for an immediate response, although this was not documented in the case record. During the debriefing with Baltimore City supervisors, it became clear that there is a tendency to over-identify the need to respond immediately (“Some days everything is an immediate”). Additionally, they indicated that sometimes an “immediate” response may not be documented in the case record but is communicated to the investigation unit supervisors verbally.

The post-implementation case reading showed that, while the quality of narrative support for some neglect items was lacking, sufficient narrative support was provided for a greater proportion of intakes post-implementation than pre-implementation. Of the 60 cases that were read from the post-implementation sample, nine were screened out, and in only one case did the narrative not support this decision. Of the 51 cases that were screened in, five cases (10%) had no allegations supported in the case narrative and five cases (10%) had some of the allegations not supported in the case narrative. Most often, the narrative did not support neglect items; again, failure to protect and risk of harm were the most prevalent.

Of the 51 screened-in referrals to which response times were assigned, 12 (24%) were assigned an immediate response time that was not supported by the case narrative. Five of the cases had responses not supported by case record because the case readers did not see the screen-in decision supported by case narrative.

Figure 1



### Summary of Findings

The findings from the pre- and post-implementation case readings suggest that the use of the SDM screening assessments improved workers' documentation within the case narrative and possibly improved the screening decisions they made. In the post-implementation case reading, more of the records included narrative that clearly supported the screen-in/screen-out decision. Additionally, of those cases that were screened in, more of the case records had documentation that supported assigned response time. Because of the relatively small sample size, these improvements cannot be identified as statistically significant. However, note that the goal of increased consistency between workers and jurisdictions regarding the screening decision will be well served with clear documentation that supports screening decisions.

## **INTER-RATER RELIABILITY TESTING**

The purpose of the screening assessment is to guide workers to a screening decision based on departmental policies and existing definitions of maltreatment. A sample of workers who field-tested the screening assessment also participated in inter-rater reliability testing. The purpose of this testing was to ensure that the screening assessment and associated definitions resulted in consistent decision making among various workers when presented with the same case information.

### **Methods**

A sample of 16 workers participated in inter-rater reliability testing in mid-September 2008. One worker was from Anne Arundel County, two were from Montgomery County, and the remaining were from Baltimore City. The objective of the test was to examine the consistency of workers' responses based on screening item definitions applied to the same case vignettes. All participants in the inter-rater reliability testing completed an SDM screening assessment on a total of 12 case vignettes. They completed the screening assessment using the web-based screening tool, which includes immediate access to definitions for each item.

The 12 case vignettes were based on actual case files of referrals made to Baltimore City DSS. The complete case narrative was pulled from MD CHESSIE and sanitized of identifying information. In some vignettes, client characteristics that would not impact the screening decision were changed to further disguise the referral.

CRC staff analyzed the results of the test using percent agreement, which is the percentage of times raters reached the same conclusion for each item on an assessment. As a measure, percent agreement is easy to understand and enables comparisons, but does not control for the degree to which different raters might select the same response by chance. The statistical

tests that control for chance agreement require a large number of vignettes, which is difficult in a practical setting. Thus, the primary measure for this analysis was percent agreement. A 75% threshold, the equivalent of the same response from three out of four people, can be used to assess the extent of agreement.

## **Findings**

Table 2 reviews the average rate of percent agreement among the 12 case vignettes reviewed by each of the 16 participants. Average rates of agreement were very high for the initial and final screening decisions (these decisions vary based on a number of policy-driven overrides that can be applied). Only one case vignette had screening decision agreement rates lower than 75%. This report from a third party involved a teenage girl disclosing an open-hand slap from her father with no marks visible at the time of disclosure.

Average rates of agreement were 80% or higher for individual screening items. The minimum rate of agreement was lower for some items, however, indicating that participants had lower rates of agreement on some items when assessing one or more of the case vignettes. Items with a minimum rate lower than 70% included the following:

- Caregiver action that is likely to cause injury.
- Inadequate food/nutrition, or signs of malnutrition.
- Inadequate clothing or hygiene.
- Inappropriate supervision.
- Conditions exist that create substantial likelihood that the child will be harmed due to caregiver neglect.

CRC staff and a workgroup of Maryland staff reviewed these items subsequent to testing to determine how to strengthen them. The summary section reviews these changes.



<b>Table 2</b>			
<b>Inter-rater Agreement Rates for Pilot Screening Assessment Items (Based on 12 Vignettes)</b>			
<b>Screening Assessment Item</b>	<b>Rate of Agreement</b>		
	<b>Average</b>	<b>Minimum</b>	<b>Maximum</b>
<b>Initial Decision</b>	<b>88.4%</b>	<b>66.3%</b>	<b>100.0%</b>
<b>Final Decision</b>	<b>92.5%</b>	<b>69.8%</b>	<b>100.0%</b>
<b>PHYSICAL ABUSE</b>			
Suspicious death of a child due to abuse	100.0%	100.0%	100.0%
Non-accidental physical injury	99.5%	93.8%	100.0%
Injury inconsistent with explanation	99.5%	93.8%	100.0%
Past injuries that appear suspicious	100.0%	100.0%	100.0%
Giving child toxic chemicals, alcohol, or drugs	100.0%	100.0%	100.0%
Caregiver action that is likely to cause injury	86.5%	56.2%	100.0%
<b>SEXUAL ABUSE</b>			
Sexual molestation of a child by an adult caregiver, family member, or household member	100.0%	100.0%	100.0%
Sexual act(s) among siblings or other children living in the home that is outside of normal exploration	100.0%	100.0%	100.0%
Sexual exploitation of a child by an adult caregiver, family member, or household member	100.0%	100.0%	100.0%
Physical, behavioral, or suspicious indicators consistent with sexual abuse	100.0%	100.0%	100.0%
<b>NEGLECT</b>			
Suspicious death of a child due to neglect	100.0%	100.0%	100.0%
Signs or diagnosis of non-organic failure to thrive	100.0%	100.0%	100.0%
Inadequate food/nutrition, or signs of malnutrition	95.8%	50.0%	100.0%
Exposure to unsafe conditions in the home	93.2%	75.0%	100.0%
Inadequate clothing or hygiene	96.4%	56.2%	100.0%
Inappropriate supervision	90.1%	62.5%	100.0%
<b>Failure to Protect</b>	97.4%	87.5%	100.0%
<b>Abandonment</b>			
A child of any age has been abandoned	99.0%	93.8%	100.0%
A child is being discharged from a facility and parents refuse to pick up or plan for the child	100.0%	100.0%	100.0%
<b>Unattended Child</b>			
A child of any age who is physically, intellectually, or cognitively disabled is left unsupervised or with responsibilities beyond his or her capabilities	99.0%	93.8%	100.0%
A child under the age of 8 has been left alone or in the care of an inappropriate caregiver	95.3%	81.2%	100.0%

<b>Table 2</b>			
<b>Inter-rater Agreement Rates for Pilot Screening Assessment Items (Based on 12 Vignettes)</b>			
<b>Screening Assessment Item</b>	<b>Rate of Agreement</b>		
	<b>Average</b>	<b>Minimum</b>	<b>Maximum</b>
A child over the age of 8 has been left alone without support systems for long periods of time or with responsibilities beyond his or her capabilities	97.9%	87.5%	100.0%
<b>Risk of Harm</b>			
Prior death or serious injury of a child due to child abuse or neglect, and a new child is now in the home	100.0%	100.0%	100.0%
Drug-exposed newborn	99.0%	87.5%	100.0%
Known sexual perpetrator has unsupervised or unrestricted access to child	100.0%	100.0%	100.0%
Conditions exist that create substantial likelihood that the child will be harmed due to caregiver neglect	80.2%	56.2%	100.0%
<b>Medical Neglect</b>			
The unreasonable delay, refusal, or failure on the part of the caregiver to seek, obtain, and/or maintain necessary medical, dental, or mental health care	99.0%	93.8%	100.0%

<b>Table 2</b>			
<b>Inter-rater Agreement Rates for Pilot Screening Assessment Items (Based on 12 Vignettes)</b>			
<b>Screening Assessment Item</b>	<b>Rate of Agreement</b>		
	<b>Average</b>	<b>Minimum</b>	<b>Maximum</b>
<b>MENTAL INJURY</b>			
<b>Abuse</b>			
A child has an observable, identifiable, and substantial impairment of his/her mental or psychological ability to function as a result of an act of a parent, permanent or temporary caregiver, or household or family member	100.0%	100.0%	100.0%
<b>Neglect</b>			
A child has an observable, identifiable, and substantial impairment of his/her mental or psychological ability to function as a result of failure to act by a parent, permanent or temporary caregiver, or household or family member	100.0%	100.0%	100.0%

## Summary

This test of inter-rater reliability among field staff from select offices showed that the screening assessment resulted in high levels of percent agreement. Average rates of agreement were greater than 75% for each item on the screening assessment. The screening decisions had very high rates of agreement (an average of 88–92%).

While average rates of agreement were high, participants had lower than 75% agreement on some items for one or more cases. It may be that workers were less likely to refer to definitions for these items, which resulted in lower levels of percent agreement. CRC and Maryland staff revisited these items to determine whether they could be strengthened. It may also be beneficial to address these areas of assessment and documentation specifically in subsequent trainings.

## **SUMMARY OF FINDINGS FROM THE FIELD TEST**

The administrators of Maryland's SSA recognize the importance of reliable decision making in social services, and therefore invested in a field test of a screening assessment designed to ensure the consistency of worker decisions about whether or not to screen in a report. Evaluation of the field-tested screening assessment focused on how well the screening assessment helped workers reach reliable decisions and could provide narrative support for those decisions. The field test included two evaluation components:

- Inter-rater reliability testing, which showed that the screening assessment resulted in high rates of agreement among workers presented with the same case information; and
- A pre- and post-implementation review of cases, which showed that the screening assessment improved the quality of case narratives and justifications for both the screening and the response priority decisions.

These findings suggest that the screening assessment, when accompanied by strategies to ensure successful implementation, can improve the consistency of screening decisions made as well as narrative justification for the decisions.

### **Review of Changes to Select Screening Items**

While average inter-rater agreement rates were high, participants had rates of agreement lower than 75% for some items for one or more cases. To improve the assessment prior to statewide implementation, CRC and Maryland staff reviewed these items to determine whether their definitions could be strengthened.

In addition, field test participants identified some items that differed from existing policies. After field test completion, Maryland made the following changes to the screening and response time assessment:

- *Abuse: Caregiver action that is likely to cause injury* is now *Caregiver action that likely caused injury*. Under Maryland statute, an investigation of abuse is one that involves an injury. While it is not necessary for the reporter to identify the type or extent of the injury, he/she must be reporting caregiver behavior that has caused or likely caused injury. As initially written on the SDM screening tool and then defined within the definitions, this item went beyond statutory language for abuse. The item has been rewritten and the definition changed to better match both COMAR and state policy.
- *Neglect: Inadequate food/nutrition or signs of malnutrition* remained as written on the assessment; however, the definition was changed to support acceptance of a referral that simply included allegations of significant lack of food or unmitigated hunger due to lack of food. The phrase deleted from the definition was “*lack of food has a negative impact on child’s health and welfare.*” This deletion was based on several discussions with SSA policy administrators and Baltimore City screening supervisors regarding the policy and expectation that referrals including an allegation of significant lack of food are accepted even if the child is not yet displaying a significant impact on health and welfare.
- *Neglect: Inappropriate supervision* is now *Inadequate supervision* and the definition has been shortened to focus on the lack of attention by a caregiver who is allowing the children to behave in a way that is potentially dangerous.
- *Neglect: Abandonment* was not changed, but a definition was added to this item. Although inter-rater reliability was strong for this item, it was not a primary or secondary allegation in any of the vignettes that were used in testing. It is thought that including a definition will support strong reliability during statewide implementation.
- *Neglect: Risk of Harm, Substance-exposed Newborn* was not changed on the screening tool, but a strong definition was added to include a reference of the specific circular letter that guides practice for referrals that include allegations of substance-exposed newborns.
- *Neglect: Risk of Harm, Conditions exist that create a substantial likelihood that the child will be harmed due to caregiver’s neglect* was changed to *Child’s basic needs are likely to go unmet due to caregiver impairment*. This item had among the lowest reliability ratings of any item on the assessment. Rewriting the item and definition, the focus is on child’s needs as they are impacted by caregiver impairment, and the concern that neglect will occur without intervention.

## **Review of Statewide Implementation Plan**

The purpose of the field test was to pilot the screening assessment prior to statewide implementation. Training for statewide implementation will occur in December 2008, with a short refresher training for offices who participated in the field test, followed by half-day training sessions for all screeners in the local agencies throughout Maryland. The training will focus on the importance of the screening decision, the use of the SDM screening tool, and the definitions that are included with the SDM screening tool. Trainers will include item examples and vignettes to ensure workers have practice applying the definitions.

Statewide implementation will include additional evaluation activities to measure whether implementation was successful in its goals of consistent decision making and improved screening practices. These activities will include the following:

- Pre- and post-implementation case reading to determine whether use of the screening assessment improves the quality of narrative statewide, as it did in field test sites. An additional case reading will be completed in Baltimore City to see what impact the refresher training has on the quality of case narrative and justification for screening decisions.
- Statewide inter-rater reliability testing to include 20 screeners from at least ten different jurisdictions to ensure that the screening assessment results in consistent decisions in new jurisdictions.
- A web-based survey of staff to obtain anonymous feedback about implementation of the screening assessment to determine whether additional training and supervision activities are necessary.

## REFERENCES

- Andrews, D.A., Bonta, J. & Wormith, J.S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency*, 52(1), 7–27.
- Baird, C., & Wagner, D. (2000). The relative validity of actuarial and consensus-based risk assessment systems. *Children and Youth Services Review*, 22(11/12), 839–871.
- Baird, C., Wagner, D., Healy, T., & Johnson, K. (1999). Risk assessment in child protective services: Consensus and actuarial model reliability. *Child Welfare*, 78(6), 723–748.
- Hendrix, M.X., & Rohland, B.M. (1997). Psychiatric hospitalization decision making by CMHC staff. *Community Mental Health Journal*, 33(1), 63–73.
- Munro, E. (1998). *Understanding social work: An empirical approach*. London: The Athlone Press.
- Rossi, P., Schuerman, J., & Budde, S. (1996, June). *Understanding child maltreatment decisions and those who make them*. Chicago: University of Chicago Chapin Hall Center for Children.