Descriptive and Reliability Analysis for the Community Supervision Tool (ORAS-CST)

Gary C. Mohr, Director
Ohio Department of Rehabilitation and Correction

Sara Andrews, Managing Director of Court and Community
Ohio Department of Rehabilitation and Correction

March 2014

Report and Analysis Prepared by:
Brian Kowalski, Senior Researcher
Bureau of Research and Evaluation

Descriptive and Reliability Analysis for the Community Supervision Tool (ORAS-CST)
ORAS Reliability and Quality Assurance Study

Sampling Design and Data Collection
The original sampling strategy for the APA sample was designed to collect 700 CSTs and interview assessments to be equally distributed across all 58 APA units (at about 12 CSTs per unit). Each particular CST assessment was designed to be the first CST completed on each offender for their current offense(s). Another goal of the sample design was to maximize staff variation. The data collection effort spanned about 4 months (February through May of 2013) and yielded 598 CSTs by 350 parole officers and 23 QA Evaluators for the reliability analysis.[1] The average number of cases in the reliability analysis completed by each parole officer is 1.71 (min = 1, max = 9, and s.d. = .823). The average number of cases in the reliability analysis completed by each QA evaluator is 26.00 (min = 1, max = 78, and s.d. = 25.143).

The original sampling strategy for the External Agency sample was designed to collect 700 CSTs and interview assessments across community agencies in the following categories: (1) CBCFs, (2) ISP Programs (both county and municipal), (3) Halfway Houses, and (4) PSI-Writing Probation Departments. The number of assessments per agency in this sample was based on the actual distribution of CSTs in the community (i.e., non-APA agencies). The data collection effort spanned about 4 months (February through May of 2013) and yielded 330 CSTs by 155 agency staff members and 17 QA Evaluators for the reliability analysis.[2] The average number of cases in the reliability analysis completed by each agency staff member is 2.13 (min = 1, max = 15, and s.d. = 1.815). The average number of cases in the reliability analysis completed by each QA evaluator is 19.41 (min = 1, max = 61, and s.d. = 16.908). It is important to note that QA evaluators in this study had a very challenging time even obtaining actual interviews at the external agency sites. This situation was worsened further by the large number of cases in the agency sample that were not entered into the ORAS system and the large number of cases in the agency sample that were determined to be invalid.

Reliability Analysis -- APA Sample
Please note that 32 cases (or 4.6% of the original sample collected) were removed from the reliability analysis because the parole officer failed to enter the tool into the ORAS system. Additionally, another 62 cases (or 9.0% of the original sample collected) were removed from the reliability analysis because the QA evaluator (23 cases) or the parole officer (39 cases) determined that 4 or more items were unable to be scored due to lack of information obtained in the interview (i.e., an invalid tool). As such, the reliability analysis is based on 598 CSTs (or 86.4% of the original sample collected).

Almost 30% of the individual items on the CST are found to be unreliable by conventional standards, which is 80% or higher staff agreement (see Table 1 on p. 4). This situation has slightly worsened since the pilot study where 20% of the individual items were found to be unreliable. Table 1 indicates several problematic domain areas including Peer Associations (with 3 out of 4 items found to be unreliable) and Criminal Attitudes and Behavioral Patterns (with 4 out of 7 items found to be unreliable). The unreliable items are negatively impacting the reliability of the final risk levels of the assessment tools. The interrater reliability of the final risk level is 74.4% before potential override decisions, under the 80% threshold for reliability. This reliability of the final risk level has slightly worsened since the pilot study (76.7%).

Reliability Analysis -- External Agency Sample
Please note that 69 cases (or 14.1% of the original sample collected) were removed from the reliability analysis because the agency staff member failed to enter the tool into the ORAS system. Additionally, another 92 cases (or 18.7% of the original sample collected) were removed from the reliability analysis because the QA evaluator (28 cases) or the agency staff member (64 cases) determined that 4 or more items were unable to be scored due to lack of information obtained in the interview (i.e., an invalid tool). As such, the reliability analysis is based on 330 CSTs (or 67.2% of the original sample collected).

Over 45% of the individual items on the CST are found to be unreliable by conventional standards, which is 80% or higher staff agreement (see Table 2 on p. 5). This situation has worsened since the pilot study where 20% of the

---

[1] It is important to view the number of parole officers with caution. In some cases, the staff member that completed the interview and CST assessment was not the same staff name identified in the ORAS database for various reasons.
[2] It is important to view the agency staff member numbers with caution as well. Some agencies only have one ORAS account for the entire agency and that particular employee sometimes handles all data entry tasks.

individual items were found to be unreliable. Table 2 indicates several problematic domain areas including Neighborhood Problems (with 1 out of 2 items found to be unreliable), Peer Associations (with 3 out of 4 items found to be unreliable), and Criminal Attitudes and Behavioral Patterns (with 6 out of 7 items found to be unreliable). The unreliable items are negatively impacting the reliability of the final risk levels of the assessment tools. Overall, the interrater reliability of the final risk level is 69.1% before potential override decisions, under the 80% threshold for reliability. This reliability of the final risk level has worsened since the pilot study (77.5%).

Interview Assessment
The descriptive statistics describing the interview assessments for the APA sample is based on the sample used for the reliability analysis (n = 598). The vast majority of the interviews utilized collateral information (94.0%). Only 45.0% of the interviews used a previous ORAS assessment. When a prior ORAS assessment was present, it was most likely a prior CST followed by a SRT.

112 parole officers (or 32.0% of the parole officers in the study) were recommended to be a peer model.[3] In contrast, 11 parole officers (or 3.1% of the parole officers in the study) were designated as needing an Individual Improvement Plan.[4]

The descriptive statistics describing the interview assessments for the External Agency sample is based on the sample used for the reliability analysis (n = 330). The vast majority of the interviews utilized collateral information (84.8%). Only 17.6% of the interviews used a previous ORAS assessment. When a prior ORAS assessment was present, it was most likely a prior CST.

44 agency staff members (or 28.4% of the agency staff members in the study) were recommended to be a peer model.[5] In contrast, 4 agency staff members (or 2.6% of the agency staff members in the study) were designated as needing an Individual Improvement Plan.[6]

Table 3 (on p. 6) displays descriptive statistics of interview skills for the group of interviews with reliable final risk scores, and the group of interviews with unreliable final risk scores for both the APA sample and the External Agency sample. Table 4 (on p. 7) compares the interview skills of the entire CST reliability analysis sample (both APA and External Agency samples combined) versus the cases that were dropped from the reliability analysis because the assessments were not entered into the ORAS system or the tools were invalid.

Summary and Implications
It is important to note that these findings do not diminish the importance of taking risk into account during criminal justice decision-making. The goal of this project was to assess the ORAS interview and assessment process along with offering suggestions for improvement.

With that said, the overall tool performs well below minimally acceptable levels of reliability for both the APA and community agencies. Like the other tools in the QA study, this situation seems to have worsened since the original pilot studies (data was originally collected in 2010 for the pilots).

The lack of agreement among the attitudinal and behavioral items and peer association items again stress the need for training to focus on problematic domains and unreliable items in particular. As we have observed across all tools in this study, these attitudinal and behavioral items pose threats to assessment reliability.

The findings indicate that missing information rendered the highest number of cases unusable for the CST among all the tools examined in the QA study. In fact, 21.6% of all cases collected in the combined CST sample were dropped from the analysis as the result of a missing assessment in the ORAS system (8.5% of combined sample) or an invalid assessment tool (13.0% of combined sample). Table 4 does offer some support that poor interviewing skills played a role in these invalid cases (i.e., skipping 4 or more items on the tool).

---

[3] Again, please interpret this descriptive statistic with caution (see footnote #1).
[4] Again, please see footnote #1.
[5] Again, please interpret this descriptive statistic with caution (see footnote #2).
[6] Again, please see footnote #2.

Simply passing a test to recertify is not a sufficient way to ensure consistent coding across the many possible interview scenarios that staff face daily. Improvement in these poorly scored areas should lead to greater reliability among individual items, domains, and final risk levels, which obviously has contact, resource, and safety ramifications. However, as we have also discussed, some of these items may have inherent limitations and should be the focus of future revalidation efforts.

Table 1. ORAS-CST Interrater Reliability Analysis of Scoreable Items for APA Sample. *

Items	Percent Agreement [#]
Criminal History	
1.1 Most Serious Arrest Under Age 18	**78.9**
1.2 Number of Prior Adult Felony Convictions	83.6
1.3 Prior Sentence as an Adult to a Jail or Secure Correctional Facility	81.3
1.4 Received Official Misconduct while Incarcerated as an Adult	86.5
1.5 Prior Sentence to Community Supervision as an Adult	89.1
1.6 Community Supervision Ever Been Revoked for Technical Violation as an Adult	85.6
Education, Employment, and Financial Situation	
2.1 Highest Education	91.8
2.2 Ever Suspended or Expelled from School	86.0
2.3 Employed at the Time of Arrest	81.8
2.4 Currently Employed/School	82.1
2.5 Better Use of Time	**78.6**
2.6 Current Financial Situation	80.8
Family and Social Support	
3.1 Parents have Criminal Record	89.8
3.2 Satisfied with Current Marital or Equivalent Situation	90.1
3.3 Emotional and Personal Support Available from Family or Others	82.1
3.4 Level of Satisfaction with Current Level of Support from Family or Others	83.9
3.5 Stability of Residence	89.0
Neighborhood Problems	
4.1 High Crime Area	89.0
4.2 Drugs Readily Available in Neighborhood	81.3
Substance Use	
5.1 Age First Began Regularly Using Alcohol	85.6
5.2 Most Recent Period of Abstinence from Alcohol	82.4
5.3 Ever Used Illegal Drugs	94.3
5.4 Drug Use Caused Problems	**74.4**
5.5 Drug Use Caused Problems with Employment	88.5
Peer Associations	
6.1 Criminal Friends	**67.1**
6.2 Contact with Past Criminal Peers	**56.2**
6.3 Gang Membership	95.3
6.4 Criminal Activities	**73.2**
Criminal Attitudes and Behavioral Patterns	
7.1 Criminal Attitudes	**55.4**
7.2 Expresses Concern about Others	**70.4**
7.3 Feels Lack of Control Over Events	80.8
7.4 Sees No Problem in Telling Lies	**77.6**
7.5 Engages in Risk Taking Behavior	**70.9**
7.6 Walks Away from a Fight	80.8
7.7 Believes in "Do Unto Others Before They Do Unto You"	91.8
Final Risk Level [^]	**74.4**

* ORAS Reliability and Quality Assurance Study -- APA Sample (n = 598). Please note that 32 cases were removed from the reliability analysis because the CST assessment was not entered into the database. Please also note that an additional 62 cases were removed from the reliability analysis because the evaluator or staff member determined that 4 or more items were unable to be scored due to lack of information obtained in the interview.

[#] Unreliable items that fall under the 80% threshold are indicated in bold.

[^] The final risk level is before potential override decisions.

Table 2. ORAS-CST Interrater Reliability Analysis of Scoreable Items for External Agency Sample. *

Items	Percent Agreement [#]
Criminal History	
1.1 Most Serious Arrest Under Age 18	**79.1**
1.2 Number of Prior Adult Felony Convictions	87.3
1.3 Prior Sentence as an Adult to a Jail or Secure Correctional Facility	86.4
1.4 Received Official Misconduct while Incarcerated as an Adult	91.8
1.5 Prior Sentence to Community Supervision as an Adult	93.3
1.6 Community Supervision Ever Been Revoked for Technical Violation as an Adult	**79.4**
Education, Employment, and Financial Situation	
2.1 Highest Education	95.8
2.2 Ever Suspended or Expelled from School	91.2
2.3 Employed at the Time of Arrest	83.6
2.4 Currently Employed/School	87.3
2.5 Better Use of Time	**78.8**
2.6 Current Financial Situation	**77.3**
Family and Social Support	
3.1 Parents have Criminal Record	92.1
3.2 Satisfied with Current Marital or Equivalent Situation	87.9
3.3 Emotional and Personal Support Available from Family or Others	81.5
3.4 Level of Satisfaction with Current Level of Support from Family or Others	86.7
3.5 Stability of Residence	80.6
Neighborhood Problems	
4.1 High Crime Area	83.0
4.2 Drugs Readily Available in Neighborhood	**73.3**
Substance Use	
5.1 Age First Began Regularly Using Alcohol	85.2
5.2 Most Recent Period of Abstinence from Alcohol	**77.3**
5.3 Ever Used Illegal Drugs	98.8
5.4 Drug Use Caused Problems	**73.3**
5.5 Drug Use Caused Problems with Employment	83.0
Peer Associations	
6.1 Criminal Friends	**58.5**
6.2 Contact with Past Criminal Peers	**58.2**
6.3 Gang Membership	95.8
6.4 Criminal Activities	**62.7**
Criminal Attitudes and Behavioral Patterns	
7.1 Criminal Attitudes	**54.5**
7.2 Expresses Concern about Others	**72.1**
7.3 Feels Lack of Control Over Events	**67.6**
7.4 Sees No Problem in Telling Lies	**69.1**
7.5 Engages in Risk Taking Behavior	**57.3**
7.6 Walks Away from a Fight	**76.1**
7.7 Believes in "Do Unto Others Before They Do Unto You"	86.1
Final Risk Level [^]	**69.1**

* ORAS Reliability and Quality Assurance Study -- External Agency Sample (n = 330). Please note that 69 cases were removed from the reliability analysis because the CST assessment was not entered into the database. Please also note that an additional 92 cases were removed from the reliability analysis because the evaluator or staff member determined that 4 or more items were unable to be scored due to lack of information obtained in the interview.

[#] Unreliable items that fall under the 80% threshold are indicated in bold.

[^] The final risk level is before potential override decisions.

Table 3. ORAS-CST Assessment of Interviewer Skills *

Interviewer Skills	APA Sample						External Agency Sample					
	Reliable Final Risk Level			Unreliable Final Risk Level			Reliable Final Risk Level			Unreliable Final Risk Level		
	Needs Improve	Meets Reqs	Exceeds Reqs	Needs Improve	Meets Reqs	Exceeds Reqs	Needs Improve	Meets Reqs	Exceeds Reqs	Needs Improve	Meets Reqs	Exceeds Reqs
Purpose of the interview was clearly explained	8.8	55.7	35.5	7.2	62.7	30.1	9.6	50.9	39.5	6.9	49.0	44.1
Avoided jumping to conclusions	1.8	42.2	56.0	1.3	44.4	54.2	1.3	37.3	61.4	0.0	33.3	66.7
Avoided barriers to listening	9.7	35.7	54.6	6.5	46.4	47.1	4.4	28.1	67.5	2.9	25.5	71.6
Evidence of collateral information being used [#]	8.8	43.4	45.4	11.1	45.8	37.3	7.9	41.7	39.0	6.9	45.1	41.2
Used open ended questions effectively	5.4	51.7	42.9	3.9	56.9	39.2	2.6	50.9	46.5	1.0	51.0	48.0
Used reflective statements to summarize answers	2.9	60.0	37.1	3.3	63.4	33.3	1.8	64.0	34.2	4.9	66.7	28.4
Avoided biased or leading questions	4.9	51.2	43.8	5.2	54.2	40.5	1.8	48.2	50.0	2.0	47.1	51.0
Follow up questions were used	4.0	55.5	40.4	4.6	56.9	38.6	5.7	46.9	47.4	2.0	52.9	45.1
Avoided roadblocks to motivation	2.2	44.0	53.7	5.9	43.1	51.0	2.6	32.5	64.9	1.0	31.4	67.6
Interview guide was used	4.5	58.7	36.9	2.6	60.8	36.6	5.7	44.3	50.0	2.9	56.9	40.2
Appropriate note taking didn't hinder conversation	0.4	60.0	39.6	0.7	69.9	29.4	3.5	52.6	43.9	2.9	47.1	50.0
Obtained information to score each domain area	3.4	69.4	27.2	3.9	65.4	30.7	3.5	63.6	32.9	2.9	65.7	31.4

* Please note that the values are expressed as percentages.

[#] Some cases in the category have scores of not applicable. These percentages are not shown here.

Table 4. ORAS-CST Assessment of Interviewer Skills *

Interviewer Skills	Reliability Analysis Combined			Missing Sample Cases Combined		
	Needs Improve	Meets Reqs	Exceeds Reqs	Needs Improve	Meets Reqs	Exceeds Reqs
Purpose of the interview was clearly explained	8.5	55.0	36.5	10.4	66.2	23.4
Avoided jumping to conclusions	1.4	40.4	58.2	7.1	42.9	50.0
Avoided barriers to listening	7.1	34.5	58.4	5.8	31.2	63.0
Evidence of collateral information being used [#]	8.7	43.5	42.0	21.4	46.1	28.6
Used open ended questions effectively	4.0	52.3	43.8	12.3	62.3	25.3
Used reflective statements to summarize answers	2.9	62.3	34.8	11.7	63.6	24.7
Avoided biased or leading questions	3.9	50.5	45.6	7.1	56.5	36.4
Follow up questions were used	4.3	53.3	42.3	24.0	53.2	22.7
Avoided roadblocks to motivation	2.8	39.7	57.5	3.2	41.6	55.2
Interview guide was used	4.3	55.3	40.4	20.1	54.5	25.3
Appropriate note taking did not hinder the conversation	1.5	58.4	40.1	3.9	59.1	37.0
Obtained information to score each domain area	3.4	66.9	29.6	31.2	53.9	14.9

* Reliability Analysis Combined (n = 928) and Missing Sample Cases Combined (n = 154). Please note that the values are expressed as percentages.

[#] Some cases in the category have scores of not applicable. These percentages are not shown here.