

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

Ohio Department of Rehabilitation and Correction

March 2014

# **ORAS Reliability and Quality Assurance Study**

This analysis is part of large-scale study examining the interview and assessment process for a host of ORAS instruments across four months in early 2013.

In particular, the broader study collected information on the Community Supervision Tool (CST), Prison Intake Tool (PIT), and Reentry Tool (RT).

Today, we will exclusively focus on the CST study.

# ORAS Reliability and Quality Assurance Study

The primary goals of the broader study are to (1) establish the **reliability** of each individual item and final risk level on each of the ORAS tools, and (2) assess the **interview skills** of staff members conducting the assessment process.

In designing our samples, we also attempted to maximize staff variation and maximize work site locations (i.e., prisons, APA units, and community agencies).

# ORAS Reliability and Quality Assurance Study

When we speak of reliability in the context of this study, we are specifically talking about **inter-rater reliability** which measures the **consistency** among staff members in scoring particular assessment items.

# **ORAS Reliability and Quality Assurance Study**

The staff members assessing reliability and interview skills were all certified ORAS Trainers.

These trainers were certified by the University of Cincinnati and worked for either the Ohio Department of Rehabilitation and Correction or various community agencies.

# ORAS Reliability and Quality Assurance Study

## Reliability

In order to establish the reliability of each item (and final risk level) on each of the ORAS tools, all offenders participating in the study are “double coded.”

This means that both the staff member and the evaluator complete an assessment tool (separately) for each offender from the same sources of information (i.e., the staff member’s interview and any collateral information).

# ORAS Reliability and Quality Assurance Study

## Reliability

Reliability is later captured by the percent agreement for each scoreable item and the final risk level.

Prior literature on objective prison classification generally considers items to be reliable if at least **80 percent** of the cases are in agreement (see Austin and Hardyman, 2004). We use this threshold to establish a minimum level of reliability.

# ORAS Reliability and Quality Assurance Study

## Interview Skills Assessment

The certified ORAS Trainers evaluated the interview skills of the staff members using an interview observational tool.

The evaluators were trained on the observational tool and were given a scoring guide for aid in the interview assessment.

# ORAS Reliability and Quality Assurance Study

## Interview Skills Assessment

Some examples of some of the interview observational tool items are listed below:

### Interview Assessment

Please assess interview skills using the following scale:

1 = Needs improvement

2 = Meets requirements

3 = Exceeds requirements

NA = Not applicable

Please circle one choice for each item:

Avoided jumping to conclusions <i>Comments:</i>	1	2	3	NA
Avoided barriers to listening <i>Comments:</i>	1	2	3	NA
Used open ended questions effectively <i>Comments:</i>	1	2	3	NA
Avoided biased or leading questions <i>Comments:</i>	1	2	3	NA
Appropriate note taking which did not hinder the conversation <i>Comments:</i>	1	2	3	NA

# APA Sample and Data Collection

The APA sample was designed to collect 700 CSTs and interview assessments across all 58 APA units.

The total number of assessments eventually collected by evaluators in the study is 692 observations.

# APA Sample and Data Collection

The reliability analysis for APA sample is based on 598 CSTs (or 86.4% of the original sample collected).

Within the original sample:

\*Please note that 32 cases (or 4.6%) were removed from the reliability analysis because the CST assessment was not entered into the ORAS database.

\*An additional 62 cases (or 9.0%) 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 (i.e., an invalid tool).

# **External Agency Sample and Data Collection**

The external agency sample was designed to collect 700 CSTs and interview assessments across (1) CBCFs, (2) ISP Programs (both county and municipal), (3) Halfway Houses, and (4) PSI-Writing Probation Departments.

The total number of assessments eventually collected by evaluators in the study is 491 observations.

As these numbers indicate, evaluators had difficulty even obtaining actual interviews at external agency sites.

# External Agency Sample and Data Collection

Reliability analysis for the external agency sample is based on 330 CSTs (or 67.2% of the original sample collected).

Within the original sample:

\*Please note that 69 cases (or 14.1%) were removed from the reliability analysis because the CST assessment was not entered into the ORAS database.

\*An additional 92 cases (or 18.7%) 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 (i.e., an invalid tool).

## Reliability Results

Almost **30% of the individual items** on the CST are found to be unreliable by conventional standards (i.e., 80% or higher staff agreement) for the **APA sample**.

Almost **45% of the individual items** on the CST are found to be unreliable by conventional standards for the **external agency sample**.

This situation has slightly worsened since the pilot study where 20% of the individual items were found to be unreliable in a limited, convenience sample of both APA units and various community agencies.

**ORAS-CST Interrater Reliability Analysis of Scoreable Items. \***

<b>Items</b>	<b>Percent Agreement #</b>	
	<b>APA</b>	<b>External Agencies</b>
<b>Criminal History</b>		
1.1 Most Serious Arrest Under Age 18	<b>78.9</b>	<b>79.1</b>
1.2 Number of Prior Adult Felony Convictions	83.6	87.3
1.3 Prior Sentence as an Adult to a Jail or Secure Correctional Facility	81.3	86.4
1.4 Received Official Misconduct while Incarcerated as an Adult	86.5	91.8
1.5 Prior Sentence to Community Supervision as an Adult	89.1	93.3
1.6 Community Supervision Ever Been Revoked for Technical Violation as an Adult	85.6	<b>79.4</b>
<b>Education, Employment, and Financial Situation</b>		
2.1 Highest Education	91.8	95.8
2.2 Ever Suspended or Expelled from School	86.0	91.2
2.3 Employed at the Time of Arrest	81.8	83.6
2.4 Currently Employed/School	82.1	87.3
2.5 Better Use of Time	<b>78.6</b>	<b>78.8</b>
2.6 Current Financial Situation	80.8	<b>77.3</b>

\* ORAS Reliability and Quality Assurance Study -- APA Sample (n = 598) and External Agency Sample (n = 330).

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

^ The final risk level is before potential override decisions.

**ORAS-CST Interrater Reliability Analysis of Scoreable Items. \***

<b>Items</b>	<b>Percent Agreement #</b>	
	<b>APA</b>	<b>External Agencies</b>
<b>Family and Social Support</b>		
3.1 Parents have Criminal Record	89.8	92.1
3.2 Satisfied with Current Marital or Equivalent Situation	90.1	87.9
3.3 Emotional and Personal Support Available from Family or Others	82.1	81.5
3.4 Level of Satisfaction with Current Level of Support from Family or Others	83.9	86.7
3.5 Stability of Residence	89.0	80.6
<b>Neighborhood Problems</b>		
4.1 High Crime Area	89.0	83.0
4.2 Drugs Readily Available in Neighborhood	81.3	<b>73.3</b>
<b>Substance Use</b>		
5.1 Age First Began Regularly Using Alcohol	85.6	85.2
5.2 Most Recent Period of Abstinence from Alcohol	82.4	<b>77.3</b>
5.3 Ever Used Illegal Drugs	94.3	98.8
5.4 Drug Use Caused Problems	<b>74.4</b>	<b>73.3</b>
5.5 Drug Use Caused Problems with Employment	88.5	83.0

\* ORAS Reliability and Quality Assurance Study -- APA Sample (n = 598) and External Agency Sample (n = 330).

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

^ The final risk level is before potential override decisions.

**ORAS-CST Interrater Reliability Analysis of Scoreable Items. \***

<b>Items</b>	<b>Percent Agreement #</b>	
	<b>APA</b>	<b>External Agencies</b>
<b>Peer Associations</b>		
6.1 Criminal Friends	<b>67.1</b>	<b>58.5</b>
6.2 Contact with Past Criminal Peers	<b>56.2</b>	<b>58.2</b>
6.3 Gang Membership	95.3	95.8
6.4 Criminal Activities	<b>73.2</b>	<b>62.7</b>
<b>Criminal Attitudes and Behavioral Patterns</b>		
7.1 Criminal Attitudes	<b>55.4</b>	<b>54.5</b>
7.2 Expresses Concern about Others	<b>70.4</b>	<b>72.1</b>
7.3 Feels Lack of Control Over Events	80.8	<b>67.6</b>
7.4 Sees No Problem in Telling Lies	<b>77.6</b>	<b>69.1</b>
7.5 Engages in Risk Taking Behavior	<b>70.9</b>	<b>57.3</b>
7.6 Walks Away from a Fight	80.8	<b>76.1</b>
7.7 Believes in "Do Unto Others Before They Do Unto You"	91.8	86.1
<b>Final Risk Level ^</b>	<b>74.4</b>	<b>69.1</b>

\* ORAS Reliability and Quality Assurance Study -- APA Sample (n = 598) and External Agency Sample (n = 330).

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

^ The final risk level is before potential override decisions.

# Reliability Results

For the APA sample, problem domain areas are **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).

# Reliability Results

A similar pattern exists for the external agency sample with problem 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).

# Reliability Results

The unreliable items are negatively impacting the reliability of the final risk levels of the Community Supervision Tool.

The interrater reliability of the final risk level for the **APA sample** is 74.4% before potential override decisions.

The interrater reliability of the final risk level for the **external agency sample** is 69.1% before potential override decisions.

# Interview Assessment Results

The vast majority of the interviews did utilize collateral information during the assessment process (94.0% for the APA sample and 84.8% for the external agency sample).

When a prior ORAS assessment was present, it was most likely a prior Community Supervision Tool.

# **Interview Assessment Results**

According to the results, staff members from both APA units and external agency sites overwhelmingly meet or exceeded requirements across all interviewer skill areas.

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.

# Summary and Implications

The goal of this project was to assess the ORAS interview and assessment process along with offering suggestions for improvement.

It is important to note that unfavorable reliability findings do not diminish the importance of taking risk into account during criminal justice decision-making.

### CST Scores and Recidivism from a Preliminary Revalidation\*

<b>Group</b>	<b>Number</b>	<b>Fail</b>	<b>%</b>
Low	2677	166	<b>6.2%</b>
Moderate (Low/Moderate)	2035	356	<b>17.5%</b>
High (Moderate)	921	283	<b>30.7%</b>
Very High (High)	169	90	<b>53.3%</b>
<b>Total</b>	<b>5802</b>	<b>895</b>	<b>15.4%</b>

\* 2011 Cases (one year follow-up). Female risk categories in parentheses where appropriate.

# Summary and Implications

The overall tool performs well below acceptable levels of reliability for both the APA units and community agencies in this study.

Like the other tools in the broader QA study, this situation seems to have worsened since the pilot studies (where the data was originally collected in 2010).

# Summary and Implications

Results indicate that staff members are generally conducting average to exceptional interviews (in terms of interviewing skills), but we are still observing problems with overall and item reliability.

We offer three potential solutions moving forward to address this issue and to support continuous quality improvement efforts at the departmental and agency levels.

# Summary and Implications

## Staff Training Improvement

Results indicate that attitudinal and behavioral items and peer association items especially pose interviewing challenges for staff.

\*Passing a test to “recertify” is not a sufficient way to ensure consistent coding across the many possible interview scenarios that staff face daily.

\*Greater and targeted attention in the staff training curriculum needs to focus on scoring these problem domains and unreliable items in particular.

# Summary and Implications

## ORAS Trainer Training Improvement

Anecdotal evidence from staff members suggests that the instructions on scoring particular items are different depending on which ORAS Trainer is conducting ORAS training.

ORAS Master Trainers (who are dedicated to train certified ORAS trainers) need to ensure that a consistent message concerning item scoring is being applied and taught.

# Summary and Implications

## Tool Revalidation Efforts

It is possible that some of these individual items may have inherent limitations, and as such, should be the focus of future revalidation efforts.

The University of Cincinnati has started the process of revalidating some of the ORAS tools and results are pending.

# Summary and Implications

These quality assurance efforts should lead to improvement in these poorly scored areas leading to greater reliability among individual items, domains, and final risk levels, which has offender contact, agency resource, and safety ramifications.