

# Offender Risk & Needs Assessment Instruments: A Primer for Courts

# Offender Risk & Needs Assessment Instruments: A Primer for Courts

Pamela M. Casey  
Jennifer K. Elek  
Roger K. Warren  
Fred Cheesman  
Matt Kleiman  
Brian Ostrom



Center for Sentencing Initiatives

This project was supported by Grant No. 2009-DG-BX-K030 awarded by the Bureau of Justice Assistance. The Bureau of Justice Assistance is a component of the Office of Justice Programs, which also includes the Bureau of Justice Statistics, the National Institute of Justice, the Office of Juvenile Justice and Delinquency Prevention, the SMART Office, and the Office for Victims of Crime. Points of view or opinions in this document are those of the authors and do not represent the official position or policies of the United States Department of Justice.



©National Center for State Courts, 2014

---

# OFFENDER RISK & NEEDS ASSESSMENT INSTRUMENTS: A PRIMER FOR COURTS

## EXPERT PANEL MEMBERS

**Honorable Carl Ashley**  
Milwaukee County Circuit Court

**Stephen A. Bouch**  
Principal  
JS Bouch & Associates

**Sally Kreamer**  
Director, Fifth Judicial District  
Iowa Department of Corrections

**Edward J. Latessa, Ph.D.**  
Professor & Director  
School of Criminal Justice  
University of Cincinnati

**Geraldine Nagy, Ph.D.**  
Director  
Travis County Adult Probation

**Randy K. Otto, Ph.D.**  
Department of Mental Health Law & Florida Mental Health Institute

**Honorable Ron Reinstein**  
Judicial Consultant  
Superior Court of Arizona (Retired)

**Jacey Skinner**  
Director  
Utah Sentencing Commission

**Faye S. Taxman, Ph.D.**  
Professor  
Criminology, Law and Society Department  
George Mason University

**Gina M. Vincent, Ph.D.**  
Assistant Professor, Department of Psychiatry  
Co-Director, National Youth Screening & Assessment Project  
Center for Mental Health Services Research  
University of Massachusetts Medical School

---

---

## TABLE OF CONTENTS

Acknowledgements.....	iv
Introduction.....	1
What are risk and needs assessment instruments, and why use them? .....	4
What are some examples of risk and needs assessment instruments; how do they differ? .....	8
What are the qualities of good risk and needs assessment instruments? .....	13
What practices support sound implementation of risk and needs assessment instruments? .....	20
What are the practical considerations in selecting and using risk and needs assessment instruments? .....	27
Conclusion .....	31
Notes & References .....	32
Appendix: Risk and Needs Assessment Instrument Profiles .....	A-1

---

## ACKNOWLEDGEMENTS

The Primer benefited from the knowledge, counsel, and support of many individuals. Chief among these are the members of the Expert Panel who provided expertise and advice on a topic that continues to grow and evolve. Their combined knowledge of risk and needs assessment instruments and hands-on experience with criminal justice system reforms were significant assets to the effort.

We also thank the test developers who patiently sat through our interviews and provided documents and other references for our review. Their willingness to participate in the effort was crucial to understanding each instrument and ensuring our reviews are current.

We also had the opportunity to discuss the use of various instruments with numerous judges, probation officers, and other criminal justice stakeholders we have met on our travels to different jurisdictions. We are grateful for their insights about and experiences with the various tools which have informed the Primer's suggestions and recommendations.

Finally, we extend our appreciation to the many individuals from the Bureau of Justice Assistance, Office of Justice Programs, for their support of the project and their patience in seeing the project to completion.

---

## INTRODUCTION

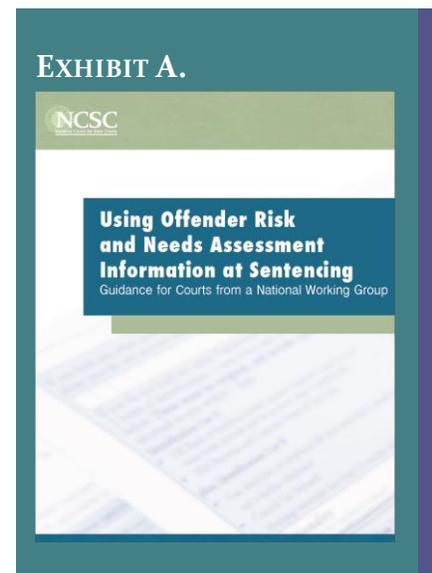
---

### *Why this Primer?*

During the last decade, the criminal justice field has focused intently on identifying programs and practices effective in reducing offender recidivism and improving public safety.<sup>1</sup> Researchers and practitioners have worked together to determine *what works best with which offenders* and, as a result, have determined that the revolving door of recidivism is not inevitable; positive outcomes for both offenders and communities are possible.<sup>2</sup> Because of the effectiveness of these evidence-based programs and practices, their use has spread to all facets of the justice system—from arrest to reentry.<sup>3</sup>

This Primer focuses on one of those decision points, *sentencing*, and on one of the tools, *risk and needs assessment (RNA) instruments*, critical to crafting sentences most likely to enhance recidivism reduction. In 2011, the National Center for State Courts published a set of guiding principles, developed by a National Working Group of practitioners and researchers, for using offender RNA information to inform sentencing decisions (Exhibit A).<sup>4</sup> The report discusses why the information is critical to the sentencing decision, how the information should be used to inform sentencing decisions, and suggestions for effectively incorporating RNA information into the sentencing process.<sup>5</sup> The guiding principles subsequently were endorsed by the Conference of Chief Justices and the Conference of State Court Administrators in a resolution, acknowledging that “research has demonstrated that the use of validated and reliable offender risk and needs assessment information to inform supervision and treatment decisions is a critical component of effective strategies to reduce recidivism.”<sup>6</sup> Specifically, the Conferences resolved to:

- “Support the National Working Group’s recommendation that offender risk and needs assessment information be available to inform judicial decisions regarding effective management and reduction of the risk of offender recidivism; and
- Endorse the guiding principles described in the National Working Group’s report as a valuable tool for state courts in crafting policies and practices to incorporate offender risk and needs assessment information in the sentencing process; and
- Encourage state and local courts to review the guiding principles and work with their justice system partners to incorporate risk and needs assessment information into the sentencing process.”



---

Although judges and other stakeholders increasingly see the value of having this information available in the sentencing process, they also have questions about how the assessments are produced and whether they are reliable, valid and fair. This Primer is a resource to help judges and others involved in sentencing understand and make knowledgeable decisions about the value and use of an assessment. It will discuss the attributes of assessment instruments that are appropriate for use in this context as well as practical considerations in selecting and properly using an RNA tool.<sup>7</sup>

The Primer also describes six of the most commonly used RNA instruments today. These descriptions are based on a review of the literature and interviews and correspondence with individuals involved in the development of the instruments. Additional research on RNA instruments is ongoing, and it is anticipated that, over time, new instruments will develop and existing instruments will be revised. Thus the criteria used for examining the six instruments in the Appendix also provide a starting point for examining any RNA instrument judges and others may consider using in their jurisdiction's sentencing process.

### *Scope of the Primer*

Practitioners use risk assessment information to inform decisions at various points in the criminal justice system. The Primer is written for judges, policy makers, and other practitioners interested in the use of RNA information at sentencing for the purpose of informing community corrections-related decisions regarding management and reduction of offender recidivism risk. It focuses on RNA instruments designed specifically to inform these community corrections-related decisions. These RNA instruments provide information relevant to sentencing considerations about an offender's amenability to supervision in the community, the level of supervision required to effectively manage the offender in the community, the types of treatment programs or other interventions most likely to reduce a specific offender's risk of reoffending, and the intensity of treatment which may be required to have recidivism-reduction effects. The Primer reviews RNA instruments that are designed for use with adult felony offenders and focused on general recidivism risk. All of the instruments provide information on an offender's risk level *and* risk factors that can be targeted with interventions to reduce recidivism.<sup>8</sup>

The Primer does not include information on instruments used exclusively at other criminal justice decision points such as pretrial release or parole, nor does it cover other instruments available to identify an offender's risk of certain types of recidivism such as violent or sexual offenses. It also does not review supplemental instruments designed to assess specialized issues such as substance abuse, mental illness, or trauma that may be warranted for use with some offenders. Some of the RNA instruments reviewed do provide additional information on offender risk at different points in the justice system (e.g., pretrial or reentry), specific types of recidivism risk (e.g., risk of committing a violent offense), or additional information regarding specific offender characteristics; but the Primer does not cover these specific aspects of the tools.<sup>9</sup> It was beyond the scope of the Primer to review all the tools focused on these aspects (e.g., all instruments

---

focused on the risk of committing a violent offense or all instruments focused on pretrial release decisions).

The remainder of the Primer covers the following five questions.

1. What are risk and needs assessment instruments, and why use them?
2. What are some examples of risk and needs assessment instruments; how do they differ?
3. What are the qualities of good risk and needs assessment instruments?
4. What practices support sound implementation of risk and needs assessment instruments?
5. What are some practical considerations in selecting and using risk and needs assessment instruments?

It is important to note that correctly using a validated RNA instrument is only one component of an evidence-based approach to reduce offender recidivism. Although the Primer is focused only on this component, readers should understand the larger context of this approach which includes, for example, matching supervision and treatment resources to an offender's risk factors, ensuring treatment programs use cognitive-behavioral skill building techniques, and selecting programs for offenders that are appropriate in light of specific offender characteristics such as gender and literacy.

---

## 1. WHAT ARE RISK AND NEEDS ASSESSMENT INSTRUMENTS, AND WHY USE THEM?

---

RNA instruments are actuarial-based tools used to classify offenders into levels of risk (e.g., low, medium, and high) and to identify and target interventions to address offender needs (e.g., antisocial attitudes, antisocial peer groups) generally related to recidivism. A RNA does not indicate whether a particular offender will actually recidivate; rather it identifies the “risk” or probability that the offender will recidivate. The probability is based on the extent to which an offender has characteristics like those of other offenders who have recidivated. For example, a RNA that results in a high risk classification means that the offender has characteristics like other offenders who have recidivated, and a low risk classification means the offender has characteristics like offenders who typically do not reoffend.<sup>10</sup>

The RNA informs risk management decisions regarding the level of supervision, i.e., the frequency and type of contact between the probation officer, client, and other individuals or agencies, required to increase the likelihood of compliance with probation conditions and ensure public safety. In addition, RNA information informs decisions regarding risk reduction strategies (e.g., cognitive behavioral programs, drug court, employment training and job assistance) that target an offender’s specific needs related to recidivism. This approach is similar to a doctor identifying a patient as a high risk for a heart attack based on several factors (e.g., high cholesterol, smoking, or poor diet) that have been shown, through research, to be related to heart disease. Although the individual may or may not actually have a heart attack, the doctor would be remiss to ignore the patient’s high risk level, and the doctor will target the patient’s treatment to those risk factors most dominant for the individual patient. Because it provides information about an offender’s relative recidivism risk and potential strategies for reducing the offender’s risk, RNA information is valuable to judges making determinations regarding an offender’s amenability to community supervision and conditions of probation in sentencing and revocation hearings.<sup>11</sup>

Research has shown the superiority of actuarial approaches to decision making over intuitive judgments in a variety of contexts, including recidivism risk.<sup>12</sup> One study of federal probation officers, for example, concluded that officers using a validated RNA tool made more consistent and accurate assessments of offender risk compared to those making unstructured professional judgments without the aid of the RNA tool.<sup>13</sup> Gottfredson and Moriarty offered several reasons for this: decision makers may not use information reliably, may not attend to base rates, may inappropriately weight predictive items, may weight items that are not predictive, and may be influenced by causal attributions or spurious correlations.<sup>14</sup> RNA instruments can assist decision makers in overcoming these issues.

To develop a RNA instrument, researchers typically collect data (or gain access to data already collected in an archive) from a representative sample of offenders on a large number of potential risk factors (e.g., criminal history, antisocial personality, school/work performance) that may be associated with recidivism. The researchers follow the offenders for a set period of time (e.g., 1-3 years) after the offenders’ prior offenses to determine whether the offenders recidivate. The data

---

from the sample of offenders are entered into a statistical model, and factors shown in the statistical model to have a significant relationship with recidivism constitute the final RNA instrument.<sup>15</sup> Subsequently, offenders who score high on the risk factors in the RNA instrument are classified as having a higher probability of reoffending; those who score lower on the risk factors are classified as having a lower probability of reoffending.

Several RNA instruments are based on the risk-need-responsivity (RNR) model. This model identifies three principles for addressing offender recidivism:<sup>16</sup>

- The **Risk** principle holds that supervision and treatment levels should match the offender's level of risk. That is, to reduce recidivism, low-risk offenders should receive less supervision and services, and higher-risk offenders should receive more intensive supervision and services.
- The **Need** principle maintains that treatment services should target an offender's dynamic risk factors or criminogenic needs (see Exhibit B<sup>17</sup>) to reduce an offender's probability of recidivism.
- The **Responsivity** principle contends that treatment interventions for offenders should use cognitive social learning strategies and be tailored to an individual offender's specific characteristics (e.g., cognitive abilities, gender) that affect successful program outcomes.

Bonta summarizes the benefit of using a RNA instrument that assists with implementing these principles:

The value of risk/need instruments is not limited to decisions around who should be supervised more closely or who should be kept in custody for the protection of the public. Because these instruments also sample criminogenic needs, they can be used to direct rehabilitation services in order to reduce offender risk.<sup>18</sup>

Research demonstrates that adherence to any one of the RNR principles correlates with a reduction in recidivism rate, and adherence to all three correlates with the highest reduction—26% — a significant decrease in current recidivism rates.<sup>19</sup> In addition to “contributing to public safety/avoiding further victimization by felony probationers and probation revocations,” the National Working Group on Using Risk and Needs Assessment Information at Sentencing highlighted several other advantages of incorporating offender assessment information into sentencing decisions:<sup>20</sup>

- Reducing prison admissions resulting from recidivism by felony probationers and probation revocations;
- Demystifying the sentencing decision and enhancing the process with scientifically-based decision tools;
- Focusing on offender accountability by requiring offenders to address their dynamic risk factors rather than placing them in programs that do not work and do not require much effort on their part;

- 
- Reducing social, economic, and family costs associated with inappropriate, and often counter-productive, interventions with low-risk offenders;
  - Ensuring sufficient prison beds for the most violent and serious offenders; and
  - Reducing prison spending by identifying offenders who can be safely and effectively supervised in the community rather than incarcerated.

#### EXHIBIT B: TERMS AND DEFINITIONS FROM THE RISK-NEEDS-RESPONSIVITY MODEL\*

Many researchers who study the link between risk factors and recidivism use RNR terms to describe various components of the link. Some of these terms are described below. Not all researchers agree with all terms and definitions. For this reason, the profile of each RNA instrument in the Appendix begins with a glossary of the terms used by the instrument's developer(s).

**Risk.** The likelihood that an offender will reoffend.

**Risk factors.** Characteristics of offenders statistically related to recidivism. Risk factors are often divided into:

- **Static risk factors.** Factors statistically related to recidivism that do not change or change in only one direction (e.g., age at first arrest, criminal history).
- **Dynamic risk factors.** Factors statistically related to recidivism that are changeable (e.g., antisocial attitudes, employment).

**Needs.** Problem areas for an offender. Needs are often divided into:

- **Criminogenic needs.** Problem areas generally related to recidivism (e.g., antisocial attitudes). These are areas typically targeted for treatment to reduce recidivism risk. Criminogenic needs and dynamic risk factors often are used interchangeably.
- **Noncriminogenic needs.** Problem areas that are not directly related to recidivism (e.g., homelessness, low self-esteem).

**Responsivity.** Targeting treatment programs to an offender's ability and learning style. Responsivity is often divided into:

- **General responsivity.** Using skill-based social learning and cognitive-behavioral programs that work to change behavior in general.
- **Specific responsivity.** Targeting treatment programs to specific offender characteristics (e.g., cognitive ability, gender).

\*Based on Andrews & Bonta (2006) and Bonta & Andrews (2007); see note 17.

Another advantage of using RNA tools is that they allow a jurisdiction to collect data over time to evaluate, for example, the effectiveness of various supervision and intervention strategies for offenders classified in different categories of recidivism risk. Data also can be used to identify the

---

types of needs most often presented by a jurisdiction's offender population and the types of supervision and intervention programs available or needed to address the needs. Thus RNA tools can also assist jurisdictions to continuously improve their allocation of resources to optimize outcome effectiveness.

---

## 2. WHAT ARE SOME EXAMPLES OF RISK AND NEEDS ASSESSMENT INSTRUMENTS; HOW DO THEY DIFFER?

---

The Appendix includes profiles of six commonly-used RNA tools:

1. Correctional Assessment and Intervention System (CAIS) which was based on the earlier Wisconsin Risk and Needs (WRN) instruments and the Client Management Classification (CMC) planning guide,
2. Correctional Offender Management Profile for Alternative Sanctions (COMPAS),
3. Level of Service Inventory-Revised (LSI-R) and Level of Service/Case Management Inventory (LS/CMI),
4. Offender Screening Tool (OST),
5. Ohio Risk Assessment System (ORAS), and
6. Static Risk and Offender Needs Guide (STRONG).

Each profile includes a glossary of terms used by the instrument developer(s) and sections on the instrument's history and current use, development, content, reliability and validity, and practical features such as automation, user qualifications, and quality assurance considerations.

The six tools include examples of instruments developed by an individual jurisdiction (i.e., OST), a state (i.e., WRN, ORAS, STRONG), or a national (i.e., CAIS, COMPAS) or provincial (LSI-R and LS/CMI) company or agency. All of the instruments have been used in multiple locations since their initial development.

RNA tools can vary in a number of ways. Several of these differences are important to an informed understanding about how a particular RNA tool may be appropriately used or implemented. Several key differences in their purpose and assessment approach follow.

### *Purpose*

As noted earlier, the Primer focuses on RNA tools developed to inform decisions about community-based supervision and treatment strategies for the general population of adult felony offenders. Several RNA tools include separate components designed for use at other decision points such as pre-trial release or release from prison (e.g., ORAS; COMPAS). Because different types of questions and outcomes are relevant for different decision points, it is important to use

### VARIATIONS IN PURPOSE AND ASSESSMENT APPROACH

- **Purpose:** How was the tool developed, for which offenders, and for which types of decisions?
- **Assessment Approach:** How does the tool calculate risk and needs; what other assessment information is provided by the tool (e.g., strengths, responsivity factors); and how is the tool administered (i.e., the methods used to conduct the assessment)?

---

any RNA tool only for the types of decision(s) for which it was intended.<sup>21</sup> Substantive differences in content may reduce predictive accuracy if a specific tool is applied at decision points other than the one at which it was originally intended for use.

RNA instruments may differ in how they define recidivism. In constructing the tools reviewed in the Primer, researchers relied on different samples of real-world offender data, outcome measures of recidivism (e.g., new arrest, conviction for a new crime, technical violation, or revocation), and follow-up periods (e.g., 1-3 years following release) for tracking reoffending.<sup>22</sup> The Community Supervision Tool (CST) of the ORAS, for example, defines recidivism as any arrest for a new crime.<sup>23</sup> The instrument developers collected data on a large number of potential RNA items from a construction sample of adult community-based offenders in Ohio and tracked new arrests over the course of a 12-month follow-up period. They retained items in the ORAS CST tool if the item correlated with rearrest during the follow-up period. The creators of the STRONG, on the other hand, examined archived, historical data on offenders released from incarceration or placed on community supervision in Washington State and defined recidivism as any subsequent felony conviction within a three year follow-up period.<sup>24</sup> Other RNA instrument developers used a more inclusive definition of recidivism, including any rule-based infraction (e.g., absconsions, rules violations, arrests, or convictions).<sup>25</sup> Differences in the type of recidivism risk calculated by a RNA tool may be meaningful in establishing local policy (or when selecting a tool to match preexisting policies), and in defining measurable recidivism reduction goals.

RNA tools may reflect the jurisdiction(s) or sample(s) of offenders on which they were developed in other ways. A RNA tool may be a valid predictor of recidivism in the particular context in which it was created, but it may not generalize well to other jurisdictions because of variations in law, policy, or the composition of the local population of adult probationers.<sup>26</sup> When one risk assessment tool originally developed in the Midwest was adopted without modification for use with probationers in New York City, researchers found that several items in the risk assessment were not related to recidivism in the New York sample.<sup>27</sup> An existing RNA tool may therefore not meet the needs of a new jurisdiction if variations in the nature or composition of the jurisdiction's target offender population alter the degree to which the instrument items and recidivism are related.<sup>28</sup>

For the above reasons, the purpose for which a RNA tool was originally designed, including the definition of recidivism used and the population on which it was developed, is an important consideration for those who use an existing RNA tool in their own jurisdiction. Subsequent validation research, if available, may also help to show that a particular RNA tool may be effectively used in a different setting or in a jurisdiction with a different demographic composition of offenders or offense types. If additional research on a particular RNA tool is not available, a good practice is to validate the instrument on the local offender population prior to adoption or full-scale implementation.<sup>29</sup>

---

## *Assessment Approach*

The researchers who created the RNA tools described in the Appendix ascribe to different theoretical approaches and different approaches to measurement. Some of the main differences among the RNA tools include (1) how they assess offender risk and needs, (2) the types of other information incorporated into the assessment, and (3) how they are administered.

**Assessment of risk and needs.** Some tools assess risk and needs together, using a single instrument and produce a composite risk and needs score, others use a single instrument and produce separate risk and needs scores, and others use separate risk and needs instruments and produce separate risk and needs scores.<sup>30</sup>

Proponents of instruments that produce a composite risk and needs score argue that all of the items in these instruments are criminogenic, i.e., they have a direct, empirically demonstrated relationship with recidivism.<sup>31</sup> In addition, because these instruments include a large proportion of items that are dynamic (i.e., changeable over time such as antisocial attitudes) as opposed to predictors that are static in nature (i.e., cannot be changed through intervention such as age), they are helpful in guiding case planning.<sup>32</sup> Assuming an instrument has been properly validated, it can help identify an offender's dynamic risk factors that, when effectively addressed, reduce recidivism risk.<sup>33</sup>

Critics of the composite score approach question the extent to which some of the dynamic risk items used in calculating the composite risk and needs score correlate with recidivism given the results of studies in different jurisdictions.<sup>34</sup> They contend that greater predictive accuracy can be achieved with shorter, more parsimonious risk scales and that separating risk and needs scales produces better measures of both.<sup>35</sup> In particular, they argue that the separate risk score is not diluted by needs items that may actually reduce the predictive ability of the risk tool. Instruments that produce a separate risk score generally rely on a smaller number (typically a dozen or less) of items found to be most predictive of recidivism in a construction sample of offenders. The separate needs score usually is based on a larger number of static and/or dynamic items that may be related to recidivism and/or identified as important by correctional officers for case management purposes.<sup>36</sup>

Critics of keeping risk and needs scores separate argue that the needs assessment portion of these RNA systems is not always subject to the same validation efforts as the risk portion.<sup>37</sup> The validated risk score is helpful in classifying an offender's risk level, but it is not helpful in identifying strategies to reduce recidivism.<sup>38</sup> Because some of the needs items may or may not be related to recidivism (e.g., items suggested by stakeholder groups as important for case planning), validation of the needs assessment is necessary to determine its effectiveness in identifying risk factors to target for intervention.

These criticisms indicate the importance for jurisdictions to look for evidence that a tool's risk and needs scores, whether provided in a composite form or separately, classify an offender correctly as low, medium, or high risk *and* also correctly identify dynamic risk factors to target for

---

risk reduction interventions.<sup>39</sup> They also provide another reason for validating any RNA instrument a jurisdiction chooses to use. Developers of RNA tools with both composite risk and needs scores, and separate risk-only scores have published research on the construction and validation of their instruments to show that each item retained in the tool has been found to be statistically related to recidivism in local construction and/or validation samples.<sup>40</sup> However, validation will ensure that the instrument retains its predictive ability when implemented in a new jurisdiction. For those instruments that provide separate needs scores, these, too, must be validated if they are to be used for identifying targets for risk reduction.<sup>41</sup>

**Other components of the assessment tool.** RNA tools also differ in the extent to which they assess other components beyond risk and needs. Some RNA tools incorporate offender strengths, also referred to as protective factors, into the assessment. A protective factor “is a variable that interacts with a risk factor to decrease the potential harmful effect of the risk factor... [acting] as a buffer that reduces the link between risk factors and later offending.”<sup>42</sup> Protective factors may include education level, employment, and the quality of family and marital relationships.<sup>43</sup> Other RNA tools include offender “responsivity factors” in the assessment. Responsivity factors are non-criminogenic offender characteristics that may affect treatment effectiveness. Responsivity factors such as the offender’s physical and mental health status, motivation to change, and learning style may affect the offender’s ability or willingness to participate in sustained treatment, or likelihood of succeeding in treatment and thus are important in case planning.<sup>44</sup>

RNA tools with separate risk and needs assessments may include both strengths (protective factors) and responsivity factors within the needs assessment. Composite RNA tools may also provide the opportunity to indicate areas of strengths (protective factors) in the full assessment (as in the LS/CMI) but separate out non-criminogenic items like responsivity factors into a different section of the tool (as in the OST).<sup>45</sup>

**Administration of the assessment.** RNA tools also differ in how they can be administered. The risk assessment component of a tool that uses separate risk and needs scales may be conducted by an intake unit using available case information and criminal records data about an offender (as with the STRONG); an interview with the offender may not be necessary.<sup>46</sup> However, the needs assessment component of such a tool and administration of composite risk-needs assessment tools both require a structured professional interview with the offender, conducted by a trained assessment administrator. Criminal records data and offender interview data may be supplemented with other methods of data collection, such as a self-report questionnaire completed by the offender undergoing assessment and/or information from collateral sources like victim statements or interviews with the offender’s family members.

Each form of assessment administration has its own pros and cons that may be weighted differently by each jurisdiction in the context of local priorities and available resources. For example, self-report surveys can be efficient, but they assume the offender understands the question being asked and also rely on the offender to supply honest answers. Structured interviews by trained professionals collect information from the offender, but in a more dynamic

---

fashion that can allow for confirmation of understanding, opportunities to probe for additional information, and a professional appraisal of the veracity of responses. Some approaches to the data review method of administration can be quite efficient: Software programs may be developed at additional cost to automate the scoring process, linking an existing data source (such as the jurisdiction's case management system) with a risk assessment application. Stakeholders should, however, be aware of the limitations of the data source upon which the risk assessment relies. For example, criminal records found in one case management system may provide only a partial picture of the offender's criminal history due to jurisdictional limitations. Stakeholders should understand the strengths and weaknesses of each data collection method and the quality of each information source(s) used by the adopted RNA tool.

To balance out shortcomings of any particular mode of assessment and as a best practice to ensure the quality of the data entering the assessment, most RNA tools require the administrator to collect information about the offender from multiple sources. For example, to obtain information about offender needs to determine appropriate treatment resources and inform case planning, a probation officer or other qualified assessment administrator will need to conduct a structured interview with the offender. Information gathered from the structured interview may be cross-checked with and/or supplemented by information provided in an offender self-report survey, a review of available records (e.g., to confirm criminal history, place of residence, educational background), and/or interviews with family members of the offender.

---

### 3. WHAT ARE THE QUALITIES OF GOOD RISK AND NEEDS ASSESSMENT INSTRUMENTS?

---

A good RNA instrument consistently produces accurate results that are fair across the types of offenders with whom the tool will be used. That is, a good tool is reliable, valid, and unbiased. Each of these general qualities is associated with specific statistical testing procedures to help ensure that the tool meets or exceeds minimum scientific standards. A description of each quality follows.

#### *Is it Reliable?*

Does the RNA tool produce consistent results if re-administered to the same person by the same or by different test administrators?<sup>47</sup> Researchers refer to this quality as reliability. Without reliability, instrument users cannot have confidence that the tool will produce an accurate result at any given time. The instrument profiles in the Appendix describe currently available research findings on the reliability of each assessment tool.

The first form of reliability referenced above – that the assessment may be administered repeatedly and produce consistent results – is called **test-retest reliability**. This form of reliability reflects the ability of the RNA instrument to generate a similar if not identical result when administered and re-administered to the same offender under the similar circumstances (i.e., by the same test administrator, assuming that nothing significant in the offender’s life has changed, for example, as a result of treatment interventions). Usually, test-retest reliability is measured using correlation statistics which show the relationship between measurements at two different points in time. Correlations range from -1.0 to +1.0, but should approach +1.0 to establish test-retest reliability. Most studies on RNA instruments do not provide information about test-retest reliability; but in broader research, scientists generally consider reliability statistics below .40 to be poor, between .40 and .59 to be fair, .60 - .74 to be good, and .75 - 1.0 to be excellent.<sup>48</sup>

The second form of reliability referenced above – that the assessment can be administered effectively by multiple test administrators – is called **inter-rater reliability** (also called inter-rater agreement). This form of reliability determines the degree to which different test administrators give the same offender similar scores on individual items as well as for the tool overall. Inter-rater reliability between two test administrators is the most common form of

#### RNA INSTRUMENT QUALITY: SIX KEY QUESTIONS

1. Is the tool reliable?
2. Is the tool valid overall?
3. Is the tool valid with all subpopulations of local offenders?
4. Is the tool easily susceptible to manipulation?
5. Has the tool been independently evaluated?
6. What are the limitations in what is empirically known about the tool?

---

reliability reported in RNA research, typically using correlation statistics.<sup>49</sup> Again, correlations may range from -1.0 to +1.0, but values should approach +1.0 to establish inter-rater reliability.

Although existing research may establish the reliability of a particular RNA tool, this information only shows that it is *possible* for the RNA tool, as it has been developed, to produce consistent results. When a tool is implemented locally, the degree of reliability *in that jurisdiction* may differ from the degree of reliability reported in prior research studies. This is because the reliability of RNA tools depends heavily on the level of training and skills of local test administrators. Both forms of reliability will be higher when test administrators receive effective, comprehensive, on-going training on how to properly use the RNA tool. Effective training will ensure that all test administrators understand the provided criteria in the same way and have the skills necessary to consistently implement established procedures when scoring the tool. Ongoing training will also help to minimize *drift* – a common tendency among test administrators to begin using the tool slightly differently from one another over time in individualistic ways that systematically distort assessment results.<sup>50</sup>

Thus when selecting and using a RNA tool, practitioners should not only be familiar with the existing research evidence demonstrating that the chosen RNA instrument is capable of achieving acceptable levels of reliability, but also understand the importance of the quality assurance mechanisms necessary to attain those levels of reliability. Those in charge of assessment should be prepared to routinely monitor reliability after the RNA tool has been implemented locally to ensure that the tool is used and continues to be used properly. This information will help determine whether the existing training package is sufficient, or if a more rigorous approach is necessary to support local use.

Reliability describes only the consistency of results generated from a RNA tool; it says nothing about how accurate those results are. Reliability is insufficient by itself to demonstrate the effectiveness of a RNA tool, but it is a necessary component of validity, which is discussed next.

### ***Is it Valid?***

The most obvious quality that a good RNA tool should have is the ability to measure what it purports to measure. This quality, called validity, focuses on measurement accuracy and also assumes that the tool can be implemented reliably (see above section).

Although validity is a singular concept, there are many different but inter-related forms of validity that reinforce one another. These multiple tests provide convergent evidence that a tool is valid. In this section, we will focus on ***predictive validity***, one of the most fundamental and important measures of validity with offender assessments.<sup>51</sup>

Predictive validity is the degree to which the results of the RNA instrument are related to behavioral outcomes of offenders *in the aggregate*. Because these testing procedures are based on averages from group data, the relationship between RNA results and behavioral outcomes for a specific individual may differ from the group results. However, group data can meaningfully

---

inform decisions about individual cases. William Grove and Paul Meehl provide the following example:

Suppose you are suffering from a distressing illness, painful or incapacitating, and your physician says that it would be a good idea to have surgeon X perform a certain radical operation in the hope of curing you. You would naturally inquire whether this operation works for this disease and how risky it is. The physician might say, "Well, it doesn't always work, but it's a pretty good operation. It does have some risk. There are people who die on the operating table, but not usually." You would ask, "Well, what percentage of times does it work? Does it work over half the time, or 90%, or what? And how many people die under the knife?"... How would you react if your physician replied, "Why are you asking me about statistics? We are talking about you – an individual patient. You are unique. Nobody is exactly like you."<sup>52</sup>

Group or aggregate data provide essential information for understanding the odds of a particular outcome. This information is applied in a number of life decisions, from more serious decisions like the medical example above to more mundane decisions like whether or not to carry an umbrella when embarking on a long walk given the local weatherman's forecast of the chance of rain. Across a number of professions and professional decision contexts, a large body of evidence demonstrates that actuarial tools produce more accurate and more reliable assessments of risk than professional judgment alone.<sup>53</sup> One of the main arguments in favor of using structured RNA tools is that, by using explicit criteria to capture information about general factors known in the scientific literature to be related to recidivism, these actuarial tools are capable of producing more consistent, accurate, objective assessments of offenders than might be generated otherwise.<sup>54</sup>

Most of the existing research on RNA instruments examines the predictive validity of the overall risk assessment component of the tool. Researchers examine the predictive validity of risk assessments empirically, using any of several different statistical techniques. The reported statistical techniques depend on the nature of the data, but at minimum will examine the relationship between the result of the assessment and a specific observed behavioral outcome (usually a form of recidivism, typically arrest or conviction for a new crime). Some of these studies also examine the extent to which each item or factor in the assessment contributes to the overall predictive validity of the risk assessment (i.e., *incremental* predictive validity).

The instrument profiles in the Appendix describe the evidence currently available on the predictive validity of each assessment tool in relation to a defined behavioral outcome (or set of outcomes). The cited evidence helps to establish the predictive validity of each tool when used under particular conditions. However, to ensure that the RNA tool is valid in a specific jurisdiction, additional local validation research is recommended. At minimum, practitioners should examine whether the tool has been validated in comparable settings with comparable target populations of offenders using the same definition of reoffending.<sup>55</sup> For a number of reasons, local validation can be helpful regardless of how often the RNA tool has been empirically validated elsewhere. Local validation research (a) will show how well the RNA tool works locally and can more concretely and convincingly demonstrate the actual benefits of using the RNA tool in that jurisdiction; (b) can help increase stakeholder confidence in the tool and encourage its

---

use; and (c) can provide invaluable research evidence to protect against potential legal challenges. Some researchers believe that local validation is required if one is seeking to adopt a RNA tool that has been validated in fewer than three similar locales.<sup>56</sup>

In the validation studies cited in the instrument profiles in the Appendix, the two most commonly reported predictive validity statistics are **correlations (*r*)** and **area-under-the-curve (AUC)** values, explained in more detail below.<sup>57</sup> Because RNA instruments classify offenders into groups of low, moderate, and high risk of recidivism to inform supervision and case planning strategies, a critical question is whether those who are classified into higher-risk groups actually show higher rates of recidivism than those classified into lower-risk groups, barring any kind of recidivism-reduction intervention. That is, an important question is not simply whether or not a risk assessment score is related to future recidivism, but whether the cutoff scores used to create the risk classification levels effectively separate low, medium, and high risk offenders.<sup>58</sup> A validation study of a good RNA tool should show the highest recidivism rates for offenders classified in the high-risk group, followed by offenders classified in the medium-risk group; the low-risk group of offenders should have the lowest recidivism rate of all.

**Correlations.** Correlations, or *r* values, are measures of association between two variables. A **point-biserial correlation ( $r_{pb}$ )** is a special kind of correlation statistic that is conducted when one of the two variables is continuous (i.e., the variable contains a range of possible values between two points, such as a risk assessment tool that generates raw scores ranging from 0 to 100), and when the other variable is dichotomous (i.e., the variable contains one of two possible values, such as when recidivism is defined as a simple yes/no to indicate whether an offender has or has not recidivated). Correlations can range from 0 to 1 (+ or -).

Correlation values provide two pieces of critical information: the direction of the relationship between two variables and the strength of that relationship. First, the sign (+ or -) indicates the direction of the relationship. In general, *r* values may be positive (“as *a* increases, *b* also increases”) or negative (“as *a* increases, *b* decreases”). All RNA tools should demonstrate an overall positive relationship with recidivism (i.e., as offender risk of recidivism scores on the RNA tool increase, actual observed recidivism should also increase). Second, the magnitude of the *r* value indicates the strength of the relationship between recidivism risk and actual recidivism. If *r* = 0, there is no relationship between recidivism risk and actual recidivism. The closer the *r* value is to 1, the stronger the relationship between the recidivism risk and actual recidivism.

Researchers will often report whether there is a “statistically significant” correlation between the raw recidivism risk scores generated by the RNA tool and offenders’ recidivistic behavior. This represents partial evidence to support a conclusion that an RNA tool does what it purports to do. However, because RNA tools are designed to produce risk level classifications, it is those classification levels – not the raw recidivism risk scores – that are actually used to inform decision-making and case planning. For this reason, better evidence of the predictive validity of a RNA tool would show that the tool accurately separates offenders into low, medium, and high risk

---

groups. A variety of statistical techniques may be used to test this, but researchers most commonly report AUC values from receiver operating characteristic analyses.

**AUC values.** AUC values represent the computed probability of the number of correct classifications, or “hits”, versus the number of incorrect classifications, or “false alarms”, by the risk assessment tool. The AUC value has advantages over other statistical techniques to help instrument users understand how well the RNA tool discriminates between offenders who will and will not reoffend, notably because it is unaffected by changes in the population’s base rate for recidivism.<sup>59</sup> An AUC = .5 means that an assessment tool is no better than chance at discriminating between recidivists and non-recidivists. The closer the AUC value is to 1, the more effective the assessment tool is at discriminating between recidivists and non-recidivists. Several groups of scientists have encouraged researchers to use and report AUCs, when possible, as the preferred measure of predictive accuracy in risk assessment, in part because the technique takes base rates into account in a standardized manner.<sup>60</sup>

When correlations and AUC values are reported as evidence for a tool’s predictive validity, researchers will interpret those values to determine how effective the tool is in practical terms. The interpretive guidelines described in Table 1 have been used by some researchers to characterize the magnitude of the “effect” of using offender risk assessment tools as small, moderate, or large.<sup>61</sup> Other researchers view these conventional guidelines as too stringent in the context of applied research and have suggested alternative cutoffs (e.g., *r* values of .1, .2, and .3 as cutoffs for small, moderate, and large effects, respectively).<sup>62</sup>

**Table 1.** General Guidelines for Interpreting Statistical Effect Sizes (Rice & Harris, 1995; 2005).

Effect	<i>r<sub>pb</sub></i>	AUC
Small	.100 to .243	.556 to .639
Moderate	.243 to .371	.639 to .714
Large	.371 or greater	.714 or greater

It is important to understand that even an effect categorized as “small” according to these conventions may meaningfully improve the assessment of risk in comparison with a business-as-usual approach.<sup>63</sup> Although scientific conventions have been established as general guidelines for interpreting the size of these effects, scientists agree that these guidelines should not be unquestioningly applied across all situations, and that “the adequacy of an assessment for a specific purpose cannot be directly inferred from single effect size indicators.”<sup>64</sup> Rather, interpreting the strength of an effect depends on a number of important factors, including but not limited to the social context of the study (e.g., what does local leadership consider to be a meaningful reduction in recidivism?) and the specific constraints of a particular research design. In fact, Rice and Harris have gone so far as to suggest that “the field of risk assessment place little reliance on plain language verbal labels because of the considerable disagreement about what they mean” among scientists, and that “clarity is best reflected by numerical characterization.”<sup>65</sup> For that reason, the Primer presents only the numerical values in the profiles of individual RNA

---

tools and does not attempt to characterize the size of the effects through a categorical label such as small, moderate, or large.

### ***Additional Validity Issues***

A good RNA tool must also produce fair results that are not systematically biased against particular subgroups of offenders and that cannot be easily manipulated by an offender to achieve desirable outcomes. Both of these concerns are subsumed within the broader research concept of validity, but merit special consideration because of the additional steps researchers must take to address these issues. Both of these issues are addressed below.

**Is it valid for all offender populations?** The instrument must produce fair and unbiased results across all of the groups of offenders on which the RNA tool will be used. This aspect of fairness is called ***differential validity***. Although the overall predictive validity of the RNA tool may have been established generally among a broad and diverse group of offenders, further examination of the predictive validity of the tool among various offender subgroups (e.g., by gender, race, ethnicity) may reveal significant differences in the degree of accuracy observed.<sup>66</sup> For example, it is possible for a risk assessment to have strong predictive validity overall, yet produce less accurate results for female offenders. Female offenders often score artificially higher (i.e., tend to be overclassified) on risk assessments that were developed with the male offender in mind and validated primarily on samples of male offenders.<sup>67</sup> Without adjustments—such as by establishing separate cutoff scores for classifying male versus female offenders as low, medium, or high risk to reoffend—tools that erroneously and systematically overclassify female offenders as higher risk will likely result in the over-supervision of female offenders in a jurisdiction that follows an evidence-based community supervision model. Moreover, some scientists have criticized the use of so-called “gender-neutral” tools with female offenders more broadly, claiming that the reliance on primarily male offender data in the instrument development process results in a tool that inadequately captures the unique criminogenic needs of female offenders.<sup>68</sup> To address these types of issues, a few providers of RNA systems now offer gender-responsive supplements in addition to the original gender-neutral version (e.g., LS/CMI, COMPAS).<sup>69</sup>

There are similar concerns regarding the predictive validity of RNA instruments for different race and ethnic groups. The extent of research on this issue varies across instruments and for different race and ethnic groups. The instrument profiles in the Appendix discuss the current research available on each tool’s predictive validity across different offender groups.

**Is it susceptible to manipulation?** Offenders may be motivated to respond artificially in ways that make them look good (called ***social desirability response bias***). Instrument developers typically incorporate strategies in the assessment process that minimize the influence of socially desirable responses on assessment results. Whether information is gathered by a trained assessment administrator conducting a structured interview with an offender or via a paper and pencil self-report measure that is completed directly by the offender, the assessment administrator is typically required to corroborate disclosed information by verifying with

---

collateral sources (e.g., official records, interviews with family or friends of the offender).<sup>70</sup> Some RNA tools with a self-report component take this a step further. The COMPAS system, for example, includes additional items in the self-report component of the assessment process that comprise what tool developers refer to as the “Lie Scale”.<sup>71</sup> These additional items are used to identify offenders who may be attempting to manipulate the results of the assessment through socially desirable responses, or what they call “faking good.” Evidence of a social desirability bias on the part of the responding offender indicates that self-reported information should be interpreted with caution and will likely require additional corroboration before RNA results can be trusted.<sup>72</sup>

### ***Additional Considerations When Reviewing Research on Risk and Needs Assessment Tools***

When reviewing the available research on a particular RNA tool, practitioners should consider two additional factors.

First, practitioners should take note of who conducted the research. Most of the available research on RNA tools has been conducted by the instrument developers themselves. Practitioners should review the research literature to determine whether the tool has been ***independently evaluated***.<sup>73</sup> That is, practitioners should determine whether the RNA tool has been rigorously evaluated by researchers who are not financially or otherwise personally invested in the success of the tool and, if so, whether those research findings support or contradict conclusions drawn by the instrument developers. Instrument developers may have an inherent conflict of interest when it comes to evaluating the success of their own tool. A bias in favor of their own tool might influence their work, consciously or not, to produce findings that cannot be reliably replicated by others. Moreover, instrument developers have more intimate knowledge about how the tool should be used that may influence how it is implemented in their testing site or how the validation study is conducted in ways that the typical user or independent researcher may not be able to duplicate from documented sources. For these reasons, it is always helpful to know whether existing research descriptions about the reliability, validity, and fairness of a tool have been replicated by others.

Second, practitioners should also understand the broader limitations of what is known about a particular tool. In researching the above psychometric properties of available RNA tools, practitioners will learn that the amount and quality of empirical research conducted varies, sometimes substantially, among the different instruments. RNA tools that have been in use longer, such as the LSI-R, will—and should—have been subjected to more rigorous evaluations and meta-analyses (analyses of the results of multiple studies) and should be supported by more documented evidence of their psychometric properties. However, simply because one RNA instrument has been studied more comprehensively than another does not necessarily mean it is a more valid tool than more recent instruments. Practical considerations, such as the resources needed to support more rigorous validation, may influence a decision about whether to use a well-studied older tool or a promising newer one. Some additional practical considerations are discussed in the next section.

---

## 4. WHAT PRACTICES SUPPORT SOUND IMPLEMENTATION OF RISK AND NEEDS ASSESSMENT INSTRUMENTS?

---

Use of a validated RNA tool is a necessary but not sufficient condition to ensure effective community-based sentencing practices. Line staff also must be equipped with the knowledge and skills necessary to use the tool properly, and management must ensure that line staff administer the tool correctly and consistently over time. A rigorous quality assurance program, including initial and ongoing staff training, coaching or mentoring, routine data monitoring, and fidelity testing (i.e., ensuring that the RNA tool is administered as it was designed), should be instituted to ensure effective implementation.

This section further discusses the importance of instrument validation and quality assurance, and key considerations at each step.

### *Instrument Validation*

**Purposes of validation.** Validation is essential to demonstrate the predictive accuracy of a RNA tool. As discussed in Section 3, the RNA tool must be supported by empirical research demonstrating that it meets basic scientific accuracy requirements in the prediction of rearrest, reconviction, or other recidivism measure of interest. Any sentencing or treatment decisions based on a RNA tool which grossly misclassifies the risk levels of offenders may not simply fail to improve outcomes; they may actually do harm to the offender. For truly high-risk offenders, less intensive supervision and treatment interventions may be ineffective.<sup>74</sup> And mandating truly low-risk offenders into more intensive supervision and/or treatment services may actually increase their recidivism risk.<sup>75</sup>

Thus a jurisdiction should not implement a RNA tool without evidentiary support that the tool appropriately categorizes the types of offenders with which the tool will be used into groups exhibiting clearly distinct probabilities of recidivism.<sup>76</sup>

Instrument validation is not only important to ensure that decision making is informed by sound data, but also to establish stakeholder confidence in the RNA tool. If probation officers, judges, and other stakeholders do not trust that the tool will enhance decision-making effectiveness, they may not use or implement the tool as intended, thereby undermining the validity of the tool. In

### PRACTICES TO SUPPORT SOUND IMPLEMENTATION

- Use a validated RNA instrument, for which validity has been demonstrated generally, is established locally, and is re-established periodically.
- Provide comprehensive initial & ongoing refresher training to all stakeholders on how to properly administer the RNA tool and understand and use its results. Develop an internal capacity to train so that these practices are sustainable.
- Routinely monitor RNA administrators for fidelity regarding proper use of overrides and consistency in scoring.

---

Maricopa County, Arizona, for example, part of the impetus for developing the OST was the observation that probation officers were not implementing the prior RNA tool as instructed because they did not believe that the tool was helpful in decision-making: Probation officers completed the tool simply because it “had to be” done.<sup>77</sup> Validation studies can provide stakeholder groups with concrete empirical evidence of the instrument’s functional value with the local offender population. This information may help to secure stakeholder buy-in when introducing evidence-based policies and practices for the first time or when integrating a new RNA instrument into existing practices. Judges and other stakeholders are more likely to support institutional changes if persuasive evidence supporting those decisions is also shared.<sup>78</sup>

**Local vs. general validation.** Instrument validity may be established locally (i.e., by commissioning a validation study within the jurisdiction in which the tool will be used) or by referencing a general body of existing validation research. A review of the existing research literature will help to determine whether or not the tool has already been validated for use in similar locations or with similar types of offenders as in one’s own jurisdiction. Some of the more established and more popular RNA tools have the benefit of a long history of research on instrument validity in an array of contexts, in a number of different jurisdictions, and conducted by a number of independent researchers. In some cases, the vendor or instrument developer warehouses data from all validation studies and can reference this data bank to determine the need for local validation. Some scientists and practitioners have indicated that if the RNA tool was developed for use with a similar population and has been validated multiple times in similar settings, or, regardless of the population on which it was developed, has been validated in at least three different jurisdictions with a similar population, setting, and definition of reoffending, local validation is not required.<sup>79</sup>

Jurisdictions can still benefit greatly from validating the chosen RNA tool locally even if instrument validity has been established generally. The same scientist-practitioner group that indicated that local validation may not be necessary in certain cases also recommends that validity still be assessed locally for any RNA tool of the type reviewed in this Primer.<sup>80</sup> As previously mentioned, differences in policy, procedure, or the makeup of the offender population may alter the predictive accuracy of a RNA tool. A local validation study will (a) inform any modifications that must be made to the content of the tool to optimize predictive validity in the local jurisdiction and ensure that it meets basic minimum scientific standards, and (b) inform the development of appropriate cutoff values for categorizing offenders into different risk levels based on actual observed differences in the probability of reoffending within the local population (also called *norming*). Judges and probation officers will be reassured that they are using a scientifically supported tool appropriate for their jurisdiction that can be confidently defended as objective, valid, and reliable. In Washington State, for example, where Department of Corrections officers may be civilly liable for their case plan decisions, a tool validated statewide offers a sense of security and protection against such liability.<sup>81</sup> This information is useful as long as the nature of the statewide sample on which the tool is validated mirrors the local population on which it is used. When properly validated, stakeholders can more confidently speak to the accuracy of the

---

classification schemes in use; the RNA tool and decisions predicated on information provided by the tool will be able to withstand critical examination.

**Revalidation.** Periodic revalidation studies of the RNA tool may also be necessary, particularly following any significant changes in local law, policing, composition of the community, or other factors that could impact offense rates or alter the common types of offending over time. Recommendations vary regarding how frequently revalidation studies should take place: One RNA instrument researcher interviewed recommended conducting revalidation studies at periodic intervals of every 3-5 years, and another instrument developer indicated that the frequency of revalidation work needed may depend on the type of assessment instrument used.<sup>82</sup> RNA tools that were developed based on emerging statistical trends observed in the relationship between existing offender data (usually convenience data like criminal history and other readily available information) and recidivism, for example, may have less stable predictive validity than RNA tools which capture information on the kinds of characteristics identified in the broader research literature as associated with criminal behavior. This is because changes in the nature and rate of recidivism on which these “statistically developed” tools are predicated, and in other factors such as contemporary community supervision practices, may reduce the predictive validity of the original assessment tool over time. In a reexamination of the original Wisconsin risk assessment tool, for example, researchers found that changes to the items and weights of the original instrument and adjustments to the risk level cutoff scores were needed in order to support continued confidence in the predictive validity of the tool.<sup>83</sup> A periodic review of classification practices will help determine whether any changes or “recalibrations” to the tool are necessary to ensure continued accuracy and appropriate classification of the local offender population over time.<sup>84</sup>

### ***Implementation Quality***

In addition to ensuring scientific support for the validity of the RNA tool, a jurisdiction should install a comprehensive plan to ensure that all users implement the tool according to its design. Without assurance of implementation quality, even a good RNA tool can produce poor or, at best, inconsistent outcomes.<sup>85</sup> A rigorous quality assurance program will not only include comprehensive and sustainable training for assessment administrators and for all users of assessment information, but also include routine quality assurance monitoring and periodic fidelity (or reliability) testing of assessment results. These components are discussed below.

### **Comprehensive and sustainable training.**

- ***Initial training and internal capacity to train.*** Users of commercially available RNA tools are generally required to undergo initial training on proper usage of the tool and the associated software before they are permitted to administer the tool. For most commercially available tools, external providers typically offer a basic two- to three-day initial training package, which covers the minimum training necessary to administer the tool. These providers also offer “train the trainers” programs to allow local jurisdictions to develop the capacity to

conduct standard trainings internally, as well as specialized courses designed to boost supplemental skills (e.g., courses on motivational interviewing, effective case planning). Establishing an internal capacity to train may be helpful not only in creating a sustainable training program for instrument administrators, but also in creating a training program to educate judges, attorneys, and other stakeholders who receive RNA information. Educating stakeholders on when and how RNA information may be appropriately applied in decision-making is a critical component of implementation that should not be overlooked, as they must understand the prescribed uses and limitations of RNA information in order to apply this knowledge effectively.<sup>86</sup> See the instrument profiles in the Appendix for details on training requirements and packages for each instrument.

If adopting or using a non-proprietary tool without an established or prepackaged training program, a training program will need to be developed from the ground up before the RNA tool can be installed. Those charged with developing the training program to support RNA installation should be knowledgeable about training strategies that optimize skill development and increase the likelihood that trained skills will be applied in practice. In a broad synthesis of implementation research literature, some researchers cited general estimates that only about 10% of trained material is typically retained by trainees.<sup>87</sup> Behavioral change is much more likely when staff members are provided with meaningful opportunities to directly apply trained skills in practical scenarios and to obtain feedback or coaching guidance for improving performance. When theory and discussion are augmented with demonstration, practice, feedback, and on-the-job coaching, 95% of trained material is retained and put into practice (see Table 2).<sup>88</sup>

**Table 2.** Summary of a Meta-Analysis of the Effects of Training and Coaching on Teachers’ Implementation in the Classroom (Joyce & Showers, 2002; excerpted from Fixsen et al., 2005, p. 30).

Training Components	Outcomes		
	(% of participants demonstrating knowledge and new skills in a training setting, and using new skills in the classroom)		
	Knowledge	Skill Demonstration	Use in Classroom
Theory and Discussion	10%	5%	0%
+ Demonstration in Training	30%	20%	0%
+ Practice & Feedback in Training	60%	60%	5%
+ Coaching in the Classroom	95%	95%	95%

- *Other ongoing training efforts.* Periodic booster or refresher training is important to prevent a problem commonly referred to as **drift**, in which test administrators start to use the same RNA tool slightly differently from one another over time in individualistic ways that distort assessment results and reduce accuracy. To prevent drift in how the RNA tool is administered and used over time, experts recommend that staff receive **refresher** (or **booster**) **training** every six months.<sup>89</sup> Refresher training should cover assessment administration as well as guidance on interpreting the results of the RNA assessment for use in supervision and case planning. Some form of refresher training is necessary not just for assessment administrators,

---

but for all those who receive RNA information for use in decision-making (i.e., probation officers in the field, judges, attorneys). As indicated above, ongoing on-the-job coaching or mentoring strategies may help to support high-fidelity implementation.<sup>90</sup> Some jurisdictions also utilize other strategies, such as peer support meetings or case review round-table meetings, to encourage users to discuss and constructively problem-solve implementation challenges.<sup>91</sup>

### **Quality assurance monitoring.**

- *Use of administrative overrides.* Some jurisdictions have specific offense-based policies in place for supervision of particular types of offenders (e.g., sex offenders) regardless of the assessed risk level of the offender, and may refer to these blanket policies as ***policy overrides*** of the RNA results. Typically, policy-based overrides prioritize other purposes of supervision such as risk management rather than recidivism reduction. This section focuses on overrides that occur as a result of an assessment administrator's subjective decision in an individual case based on his or her own professional judgment, or ***administrative overrides***.

Most RNA tools contain a discretionary administrative override function that the assessment administrator is authorized to use to modify individual RNA results. That is, if the administrator believes that certain information about the offender is not adequately captured in the assessment and that the results should be altered to better reflect this information, the administrator may make a discretionary decision in that case to modify the offender's RNA results accordingly. Most instrument developers emphatically caution against frequent use of the administrative override function and encourage a practice in which such exceptions are made in no more than 10% of all cases (overall or per assessor).<sup>92</sup> Some instrument providers recommend a lower exception rate (e.g., 2-3%).<sup>93</sup>

To date, little research exists to document the impacts of discretionary administrative overrides on the predictive accuracy of risk assessment tools. Across a number of decision-making contexts, however, the exercise of subjective judgment by a clinician or other professional with specialized expertise in the absence of an actuarial tool or other structured decision aid, referred to as ***unstructured professional judgment***, generally produces results inferior to judgments informed by these tools because humans are simply not very good at reliably and accurately identifying and weighing the complex factors that inform risk.<sup>94</sup> Frequent use of the administrative override function in an assessment tool based on the administrator's professional judgment risks diminished assessment accuracy: Studies outside of the offender risk assessment field have demonstrated that human judgment, when used *only to amend the results of an actuarial model*, still reduced predictive accuracy compared with the unmodified actuarial results.<sup>95</sup> Similarly, one recent offender risk assessment study examined the use of the professional override function in administering the LS/CMI with a sample of sex offenders.<sup>96</sup> The study showed that administrators were much more likely to apply a discretionary override to LS/CMI results in order to increase the offender's risk level than to decrease it. Importantly, the application of administrative overrides served to decrease

---

the accuracy of the assessment overall, but especially so when overriding to increase the offender's risk level. The authors discovered that administrators intuitively based their discretionary override decisions on offender characteristics that are not truly associated with recidivism risk. Although this study was conducted on a unique sample of adult felony offenders, until more research on the use of administrative overrides with the general population of felony offenders becomes available, it highlights the potential risks inherent in the practice of overriding assessment scores and serves as a caution against frequent use of the override function.

A high rate of overrides among risk assessment administrators may be indicative of more fundamental implementation problems. It may signal, for example, that more training is required on how to properly administer the RNA tool and should trigger targeted coaching, mentoring, retraining, or other quality assurance efforts. Alternatively, liberal use of the override function may be a symptom of a different problem: that staff users have low confidence in the utility of the RNA tool. In that instance, assuming the tool has been properly validated, additional efforts to educate staff on the research supporting use of the tool may be needed. (See *Instrument Validation* in this section, above.)

To deter frequent and inappropriate use of the override function, court and probation leaders have taken different approaches. Some jurisdictions permit the use of an administrative override in exceptional circumstances only, and have established protocols requiring clear documentation of reasoning and formal approval by a supervisor.<sup>97</sup> Alternatively, other jurisdictions have elected to prohibit administrative overrides entirely.<sup>98</sup>

- *Data monitoring.* A good quality assurance program should include two main efforts. First, the jurisdiction should be able to show that as a result of training, different RNA instrument administrators are able to produce consistent scores on the RNA tool and its individual items. That is, an individual should receive the same RNA results regardless of the administrator conducting the assessment. As discussed in Section 3, this type of inter-rater reliability has significant implications for the validity and credibility of the tool. A properly validated RNA tool will be supported by evidence that it *can* be scored consistently to produce reliable results. Inter-rater reliability tests will show whether the tool is being *administered* correctly by staff in the local jurisdiction, and whether the reported results from use of the validated RNA tool can be trusted.

Second, the jurisdiction should be able to identify staff members who are using the RNA tool according to established procedure and those who may require additional training or other supportive services to build the required assessment skills. Supervisors may conduct **case audits**, a periodic review of line staff assessment and scoring practices, to ensure adherence to established protocol. Supervisors may also observe and critique samples of assessment interviews in person or on audio or video tapes to provide line staff with performance feedback.<sup>99</sup> In addition, aggregate data monitoring procedures may be helpful. Some

---

researchers suggest that jurisdictions examine data collected over time to determine, for example, whether the percentage of assessed offenders who fall into each risk category (low, moderate, high) are approximately equal; whether the distribution across risk categories differs substantially between females and males or between offenders of different racial or ethnic backgrounds (which may trigger further examination of the potential for bias in application of the RNA); and whether the proportion of overrides applied exceeds maximum limits recommended by instrument developers in cases overall or in cases supervised by any individual assessment officer.<sup>100</sup>

Some RNA service providers may offer trainings or add-ons to automated RNA systems designed to support fidelity testing. For example, the ORAS includes a feature which allows the client to draw random samples of cases for internal review, and clients may complete a certification course offered by the University of Cincinnati to develop internal capacity to conduct routine fidelity studies.<sup>101</sup> If a fidelity testing software program is not available through the RNA provider, local users should be able to export data from an automated RNA system for manual analysis. If internal capacity does not exist to analyze data for quality assurance purposes, the RNA provider or other research contractors may be available to provide research services.

---

## 5. WHAT ARE THE PRACTICAL CONSIDERATIONS IN SELECTING AND USING RISK AND NEEDS ASSESSMENT INSTRUMENTS?

---

Several practical considerations will likely inform decisions about selecting and using a RNA instrument. These include considerations related to the availability of services designed to support implementation and maintenance of the RNA tool, associated costs, and the ease of use.

### *Availability of Support Options*

Some vendors may operate as a “one-stop shop,” offering not only the RNA tool itself, but also the research and training services as described in section 4 that are necessary to support quality implementation and on-going maintenance of the system over time. Vendors may conduct validation and fidelity studies, and provide train-the-trainer and user training programs to support the use of the tool. They may also establish forums for users of the tool to submit questions to instrument developers, ask questions of their peers in the community, and share information on associated policies, procedures, and practices. Vendors may also offer a range of specialized software packages that may be tailored to the needs of the client jurisdiction. The software will, at a minimum, compute the results of the assessment and generate individual assessment reports, saving time and minimizing user error. Other software options typically bundle a case management system with the automated assessment. In addition to a case planning function, these systems enable the tracking of offender outcomes and may include a variety of customizable aggregate report generation options. The case management system may be housed by the vendor on a remote server that requires local users to have internet access and assigned user login information. Often, the software bundle may be purchased and installed on a server owned and operated by the local jurisdiction. Most vendors also offer technology solutions to integrate the RNA software bundle with a client jurisdiction’s existing case management system.

#### RNA INSTRUMENTS: PRACTICAL CONSIDERATIONS

- **Availability of Support Options:** What services (e.g., RNA and reporting software, custom IT integration, user training, train-the-trainer training, quality assurance monitoring, validation research) does the RNA vendor provide? Alternatively, what support services are not available?
- **Costs:** What costs are associated with implementation and ongoing use of the RNA tool (e.g., instrument & software subscription costs, initial & ongoing stakeholder training costs, quality assurance protocol development & monitoring costs, periodic validation research)?
- **Ease of Use:** How easy is the tool to implement, administer, and use to inform decision-making?

---

Other vendors may offer only a limited array of support services. Some may offer a large menu of support services by subcontracting with external agencies to provide the services. A vendor may house a strong team of software developers and provide sophisticated IT services directly to clients, for example, but subcontract with external consultants when validation research services are required. The in-house expertise of the identified vendor may have important implications for management needs and for ongoing costs associated with use of a particular RNA tool.

### *Costs*

Many costs associated with the use of RNA tools extend beyond the pricing of the instrument itself. Other costs include research, training, software, and other technical assistance services of various forms. For proprietary RNA tools, a batch order for a defined number of assessments may be placed, or a bulk rate may be negotiated per assessment or per case for which assessments and reassessments are to be conducted. Validation research studies and fidelity testing may be included as part of the original service agreement or may be available at an additional cost. Training services are also an additional cost and are typically priced per session. However, most vendors supply train-the-trainer programs to allow local jurisdictions to develop the in-house capacity to conduct future user training sessions. The costs of various software solutions will vary, although ongoing technical assistance support is usually complimentary.

Some RNA instruments are non-proprietary and may be available for use free of charge, but calculations of total cost should consider the availability and pricing of other important support services, such as validation research, fidelity testing, training, and customization of software packages designed for the RNA tool. Some vendors offer support services for the non-proprietary tools reviewed in the Primer's Appendix. If external support is not available or expensive, the jurisdiction should determine whether the costs associated with developing support services or processes of a comparable quality in-house are worth the savings associated with the use of a free RNA tool.

### *Ease of Use*

Finally, the jurisdiction should consider the broader ramifications of adopting a particular RNA tool. This includes considerations related to the user qualifications or requirements to administer the tool. Is the tool complex and difficult to understand? How much staff training is necessary before the tool can be used as compared with other viable options? To administer the LS/CMI, for example, the vendor requires that the staff person: (1) complete a specialized training program administered by an MHS-approved trainer, or (2) document previous completion of graduate-level or professional training on psychometric testing and measurement, or (3) be closely supervised by a test administrator who has completed an approved training program or course.<sup>102</sup>

Another consideration is the amount of staff time involved in proper administration of the tool and use of RNA information. Although the availability of RNA information offers many benefits, the administration of the RNA tool and administrative processes for use of RNA information are often more time consuming than the pre-existing approach. How long will it take to administer

---

and score the assessment? How does the use of RNA information differ from the current approach, and how will the changes in workload affect operations? Will the results be reported in a manner that is easy to incorporate into existing reporting processes, including, for example, to the court? Will the reported information be easy for all users to read, understand, and consistently use in decision-making? These workload efficiency considerations may prompt a need for organizational restructuring. Some jurisdictions, for example, have elected to create a centralized unit in the probation department that is tasked with conducting all initial offender assessments as part of a diagnostic process and producing all presentence investigation reports for the court. In these jurisdictions, supervising probation officers typically conduct subsequent reassessments if the offender is placed on probation.

Finally, the degree of staff support for the use of a RNA tool is also an important consideration. How receptive are judges, staff, and other stakeholders to adoption and implementation of the RNA tool? How committed will they be to using the tool properly and consistently? Greater buy-in from stakeholders may result in more faithful implementation.<sup>103</sup> With guidance from experts, an implementation committee comprised of leadership from local stakeholder groups can be assembled to select an appropriate tool for the jurisdiction.<sup>104</sup> This level of engagement in the initial selection and development process can help to ensure that all stakeholder perspectives are heard at the outset, and can be effective in establishing the necessary foundation of support. In some cases, it may make more sense for a jurisdiction to simply expand the use of an existing RNA tool already employed by the local probation department, if the culture surrounding the use of the RNA tool is a positive one and the tool meets the psychometric standards previously described.

#### *A Note Regarding the Decision to Develop a New Risk and Needs Assessment Tool*

In some cases, jurisdictions may elect to develop, validate, implement, and support the ongoing use of their own RNA tools. Compared with adoption of a RNA tool “off-the-shelf,” this approach requires a larger initial financial investment to support the time-consuming development efforts. The jurisdiction will need to hire professional scientific research personnel with expertise in psychometrics and experience working with criminal justice populations.<sup>105</sup> These researchers should develop a RNA instrument appropriate for use in the jurisdiction, conduct an initial validation study of the new tool, establish a training curriculum for local staff and stakeholders on the proper use of the tool, help establish local capacity to implement the training curriculum in the long term by training local trainers, and provide guidance on the future steps required to maintain the overall effectiveness of the RNA instrument and assessment process over time—including periodic revalidation studies, routine fidelity testing, and other ongoing quality assurance measures. Depending on the research design, the initial validation study of a new RNA instrument alone may take several years to complete.

Because of the time involvement and financial investment associated with developing a new tool, this option may be most advantageous for jurisdictions that already use a RNA tool as part of an established use of evidence-based practices but seek performance improvements such as

---

improved predictive validity or reliability beyond what is perceived possible by using the existing tool and process. In addition, use of a locally developed RNA tool may incur fewer ongoing costs, for example, by eliminating the costs of purchasing a proprietary assessment and by assembling other support services piecemeal, perhaps through a competitive bidding process.<sup>106</sup> Local stakeholders also may feel a greater sense of ownership of the new instrument and process that can, in the long term, stimulate greater support for and more faithful implementation of the tool.

---

## CONCLUSION

---

The proper use of validated, actuarial RNA instruments in assessing the level of risk and criminogenic needs of offenders subject to probation or community supervision is an established evidence-based practice and essential to the success of any serious recidivism reduction enterprise. In this Primer, we have sought to address key questions that judges, probation leaders, and other stakeholders may have about RNA tools in order to assist them in making knowledgeable decisions about the adoption and use of such tools. We have also provided detailed user-friendly information about six commonly used tools in community supervision agencies. Armed with the information provided in this Primer we are confident that criminal justice practitioners will be well-prepared to secure accurate, objective, and reliable risk and needs assessment information on offenders within their jurisdiction.

But even the most accurate, reliable, and fair RNA tool, properly administered by well-trained staff, will not automatically result in changing offender behavior or reducing offender recidivism. A properly validated tool and well-trained officers administering the instrument are certainly two necessary conditions for the effective use of risk and needs assessment information. But much more is also required. Probation officers, judges, and other stakeholders must also be well-trained on other aspects of evidence-based corrections practice: how to use RNA information in tailoring supervision plans and probation orders, how to motivate and effectively supervise offenders to comply with conditions of probation, how to help offenders develop the skills to sustain law-abiding behaviors, and how to most effectively respond to violations of supervision conditions. In addition, sufficient demonstrably effective treatment resources must be available in the community to address offenders' criminogenic needs. Many external providers offer training programs designed to develop and enhance probation skill sets that are critical to effective supervision, and research services to evaluate treatment programs for efficacy.

Accurate assessment is essential but wasted effort unless it leads to effective supervision and treatment. Like assessment and diagnosis in medicine, accurate assessment in corrections is only the first step in the process of developing and then implementing an effective treatment plan. But the fact remains that it is a critical first step: if the initial assessment is inaccurate, the resulting course of supervision and treatment is likely to fail. The authors hope this Primer provides judges and other stakeholders with the information they need to successfully plan and undertake this critical first step in establishing sentencing and community corrections practices that are effective in reducing offender recidivism.

---

## NOTES & REFERENCES

---

<sup>1</sup> Although the last decade has seen an intense focus on identifying programs and practices that work, i.e., evidence-based practices, many researchers and practitioners began exploring the issue much earlier. See Taxman et al. (2013): “Since the early 1990s, the risk-need-responsivity (RNR) model for correctional programming has served as a framework to promote the use of evidence-based correctional strategies” (p. 73). Taxman, F. S., Pattavina, A., Caudy, M. S., Byrne, J., & Durso, J. (2013). The empirical basis for the RNR model with an updated RNR conceptual framework. In F. S. Taxman & A. Pattavina (Eds.), *Simulation strategies to reduce recidivism* (pp. 73-111). New York: Springer. Also see Eisenberg, M., & Markley, G. (1987). Something works in community supervision. *Federal Probation*, 51, 28-32.

<sup>2</sup> See, for example, Andrews, D. A., & Bonta, J. (2006). *The Psychology of Criminal Conduct*, (4th ed.). Cincinnati: Anderson.

<sup>3</sup> See, for example, the National Institute of Justice, Office of Justice Programs, CrimeSolutions.gov website at [www.crimesolutions.gov](http://www.crimesolutions.gov). For information about how programs and practices are rated as effective, see the About CrimeSolutions.gov, Program Review and Rating from Start to Finish website page at [http://www.crimesolutions.gov/about\\_starttofinish.aspx](http://www.crimesolutions.gov/about_starttofinish.aspx). Also see the National Institute of Corrections, Evidence-Based Decision Making website at <http://nicic.gov/ebdm> for information on using evidence-based practices and programs throughout the criminal justice system.

<sup>4</sup> Casey, P. M., Warren, R. K., & Elek, J. K. (2011). *Using offender risk and needs assessment information at sentencing: Guidance for courts from a National Working Group*. Williamsburg, VA: National Center for State Courts. Retrieved from <http://www.ncsc.org/sitecore/content/microsites/csi/home/Topics/~media/Microsites/Files/CSI/RNA%20Guide%20Final.ashx>

<sup>5</sup> The guiding principles expressly state that RNA information should be used for decisions regarding risk reduction and management and not for decisions regarding the appropriate punishment for an offender.

<sup>6</sup> Conference of Chief Justices & Conference of State Court Administrators. (2011). *Resolution 7 in support of the guiding principles on using risk and needs assessment information in the sentencing process*. Williamsburg, VA: National Center for State Courts. Retrieved from <http://ccj.ncsc.org/~media/Microsites/Files/CCJ/Resolutions/o8o32o11-Support-Guiding-Principles-Using-Risk-Needs-Assessment-Information-Sentencing-Process.ashx>

<sup>7</sup> RNA instruments discussed in the Primer are also referred to as “RNA tools” because they include options for generating automated case plans and various management reports.

<sup>8</sup> Some risk assessments are used to classify offenders for various reasons but do not include a needs assessment to identify intervention targets for recidivism reduction. For example, the Virginia Criminal Sentencing Commission (VCSC) developed a risk assessment instrument to identify prison-bound offenders who were low risk to reoffend for purposes of diverting them to a non-prison alternative. See Ostrom, B. J., Kleiman, M., Cheesman, F., Hansen, R. M., & Kauder, N. B. (2002). *Offender risk assessment in Virginia: A three-stage evaluation*. Williamsburg, VA: National Center for State Courts. Retrieved from [http://www.ncsonline.org/WC/Publications/Res\\_Senten\\_RiskAssessPub.pdf](http://www.ncsonline.org/WC/Publications/Res_Senten_RiskAssessPub.pdf). The Primer does not focus on these “risk-only” instruments.

<sup>9</sup> For information on specific offender characteristics, see Andrews et al. (2004): Specific offender characteristics include factors that are not linked to recidivism in the general offender population, “but when they do occur in a given case, they may take a particularly prominent role in the

---

assessment of the offender's risk" (p. 23). Examples of specific offender characteristics are problems with compliance, a diagnosis of psychopathy or other personality disorder, problem-solving and self-management skill deficits, and poor social skills. Andrews, D. A., Bonta, J. L., & Wormith, J. S. (2004). *LS/CMI Level of Service/Case Management Inventory: An offender assessment system—user's manual*. North Tonawanda, NY: Multi-Health Systems. Also see Exhibit B "Terms and Definitions from the Risk-Needs-Responsivity Model" in Section 1 "What Are Risk and Needs Assessment Instruments, and Why Use Them?"

<sup>10</sup> See pp. 29-31 in Vincent, G. M., Guy, L. S., Grisso, T. (2012, November). *Risk assessment in juvenile justice: A guidebook for implementation*. Chicago: MacArthur Foundation. Retrieved from <http://www.modelsforchange.net/publications/346>

<sup>11</sup> Casey et al. (2011) at note 4.

<sup>12</sup> Gottfredson, S. D., & Moriarty, L. J. (2006). Clinical versus actuarial judgments in criminal justice decisions: Should one replace the other? *Federal Probation*, 70(2), 15-18. See also, Harris, P. M. (2006). What community supervision officers need to know about actuarial risk assessment and clinical judgment. *Federal Probation*, 70(2), 8-14.

<sup>13</sup> Oleson, J. C., VanBenschoten, S. W., Robinson, C. R., & Lowenkamp, C. T. (2006). Training to see risk: Measuring the accuracy of clinical and actuarial risk assessments among federal probation officers. *Federal Probation*, 75(2), 52-56 at 54.

<sup>14</sup> See Gottfredson & Moriarty (2006) at note 12, p. 15.

<sup>15</sup> This is a generalization of the process. The Appendix describes the specific development process for six commonly used instruments.

<sup>16</sup> See Andrews & Bonta (2006) at note 2, pp. 279-284. Also see, Bonta, J., & Andrews, D. A. (2007). *Risk-need-responsivity model for offender assessment and rehabilitation (PS3-1/2007-6)*. Ottawa: Public Safety Canada. A summary of the RNR model also is included in Casey, et al. (2011) at note 4.

<sup>17</sup> The terms and definitions referenced in Exhibit B are drawn from Andrews & Bonta (2006) at note 2 and Bonta & Andrews (2007) at note 16.

<sup>18</sup> Bonta, J. (2007). Offender Risk Assessment and Sentencing. *Canadian Journal of Criminology and Criminal Justice*, 49, 519-529 at 520. Also see Taxman (2006): "An actuarial-based risk screen is important to determine the degree to which offenders should be given services and resources to ameliorate criminal behavior. The type of services is determined by how the offender "scores" or presents on several criminogenic areas. Those offenders with high criminogenic needs, particularly those that are high or moderate risk, should be given services to ameliorate the criminogenic need, which should reduce the risk for recidivism" (p. 7). Taxman, F. (2006). Assessment with a flair: Offender accountability in supervision plans. *Federal Probation*, 70(2), 2-7.

<sup>19</sup> Andrews & Bonta (2006) at note 2, pp. 73-74. Also see, Andrews, D. A., & Dowden, C. (2007). The risk-need-responsivity model of assessment and human service in prevention and corrections: Crime-Prevention jurisprudence. *Canadian Journal of Criminology and Criminal Justice*, 49, 439-464.

<sup>20</sup> Casey, et al. (2011) at note 4, pp. 7-8.

<sup>21</sup> Vincent et al. (2012, November) at note 10, p. 58.

<sup>22</sup> See Section 1 "What Are Risk and Needs Instruments, and Why Use Them?" for a general description of how a RNA tool is developed; also see the instrument profiles in the Appendix for a description of each tool's specific development process.

<sup>23</sup> Latessa, E., J., Lemke, R., Makarios, M., Smith, P., & Lowenkamp, C. T. (2010). The creation and validation of the Ohio Risk Assessment System (ORAS). *Federal Probation*, 74, 16-22. Retrieved

---

from <http://www.uscourts.gov/viewer.aspx?doc=/uscourts/FederalCourts/PPS/Fedprob/2010-06/index.html>

<sup>24</sup> Barnoski, R., & Drake, E. (2007). Washington's Offender Accountability Act: Department of Corrections' static risk assessment (Doc. No. 07-03-1201). Olympia, WA: Washington State Institute for Public Policy. Retrieved from

[http://www.wsipp.wa.gov/ReportFile/977/Wsipp\\_Washingtons-Offender-Accountability-Act-Department-of-Corrections-Static-Risk-Instrument\\_Full-Report-Updated-October-2008.pdf](http://www.wsipp.wa.gov/ReportFile/977/Wsipp_Washingtons-Offender-Accountability-Act-Department-of-Corrections-Static-Risk-Instrument_Full-Report-Updated-October-2008.pdf)

<sup>25</sup> Baird, C. S., Heinz, R. C., & Bemus, B. J. (1979). *The Wisconsin case classification/staff deployment project: A two-year follow-up report*. Madison, WI: Wisconsin Division of Corrections.

<sup>26</sup> Gottfredson, S., & Moriarty, L. (2006). Statistical risk assessment: old problems and new applications. *Crime & Delinquency*, 52, 178-200. doi: 10.1177/001128705281748. See also Johnson, K. D., & Hardyman, P. L. (2004). How do you know if the risk assessment instrument works? In National Institute of Corrections (Series Ed.), *Topics in Community Corrections: Assessment Issues for Managers* (pp. 20-26). Washington, DC: National Institute of Corrections.

<sup>27</sup> Wright, K. N., Clear, T. R., & Dickson, P. (1984). Universal applicability of probation risk-assessment instruments. *Criminology*, 22, 113-134. doi: 10.1111/j.1745-9125.1984.tb00291.x

<sup>28</sup> See, also, "Is it valid for all offender populations?" in Section 3 "What Are the Qualities of Good Risk and Needs Assessment Instruments?" for a discussion about gender-responsive assessment.

<sup>29</sup> For further discussion of instrument validation, see Section 4 "What Practices Support Sound Implementation of Risk and Needs Assessment Instruments?"

<sup>30</sup> Those instruments that produce a composite risk and needs score (e.g., LSI-R, LS/CMI, ORAS, OST) also provide scores on individual needs domains (e.g., antisocial associates, antisocial attitudes, family and social support), COMPAS provides a needs score for individual domains rather than a total needs score. The CAIS instrument collects information on both risk and needs factors and provides a risk score. It does not provide a total or separate needs score; rather it identifies areas that should be addressed in an offender's case plan. The WRN uses separate instruments to provide a total risk score and a total needs score. STRONG uses separate instruments to provide a total risk score and a needs score for individual domains.

<sup>31</sup> Bonta, J. (2002). Offender risk assessment: Guidelines for selection and use. *Criminal Justice and Behavior*, 29, 355-379. doi: 10.1177/0093854802029004002. See definitions of Risk-Need-Responsivity terms in Exhibit B in Section 1 "What Are Risk and Needs Assessment Instruments, and Why Use Them?"

<sup>32</sup> B. Lovins, personal communication, February 16, 2012. Latessa et al. (2010) at note 23.

<sup>33</sup> For example, see Vose, B., Smith, P., & Cullen, F. T. (2013). Predictive validity and the impact of change in total LSI-R score on recidivism. *Criminal Justice and Behavior*, 40, 1383-1396. See also Latessa, E. J., & Lovins, B. (2010). The role of offender risk assessment: A policy maker guide. *Victims & Offenders*, 5, 203-219.

<sup>34</sup> For example, see (1) Austin, J., Coleman, D., Peyton, J., & Johnson, K. D. (2003, January). *Reliability and validity study of the LSI-R risk assessment instrument*. Washington, DC: The Institute on Crime, Justice, and Corrections at The George Washington University. (2) Barnoski, R., & Aos, S. (2003). *Washington's Offender Accountability Act: An analysis of the Department of Corrections' risk assessment* (Doc. No. 03-12-1202). Olympia, WA: Washington State Institute for Public Policy. Retrieved from <http://www.wsipp.wa.gov/Reports/03-12-1202>. (3) Caudy, M. S., Durso, J. M., & Taxman, F. S. (2013). How well do dynamic needs predict recidivism? Implications for risk assessment and risk reduction. *Journal of Criminal Justice*, 41, 458-466.

<sup>35</sup> Baird, C. (February, 2009). *A question of evidence: A critique of risk assessment models used in the justice system*. Madison, WI: National Council on Crime and Delinquency. Retrieved from

---

[http://www.nccdglobal.org/sites/default/files/publication\\_pdf/special-report-evidence.pdf](http://www.nccdglobal.org/sites/default/files/publication_pdf/special-report-evidence.pdf). See also Gottfredson & Moriarty (2006) at note 26, p. 192, warning against confusing the purposes of risk and needs assessment:

Although we focus on risk and/or needs assessment as if they can be interchanged, they are very different. We would argue that predicting who will or will not behave criminally is risk assessment, whereas using predictive methods to attempt a reduction in criminality through assignment to differential treatments is needs assessment.

<sup>36</sup> The development of the Wisconsin and STRONG needs assessments, for example, involved input from probation officers regarding factors they considered important to know in preparing case plans. See Baird et al. (1979) at note 25, pp. 12-13. STRONG information is from R. Barnoski, personal communication, April 24, 2012.

<sup>37</sup> For example, the STRONG needs assessment has not been validated. R. Barnoski, personal communication, April 24, 2012. The Wisconsin needs instrument was validated for cases management purposes—not for risk reduction purposes. Researchers and practitioners wanted a tool to better estimate the amount of supervision time a case would require based on the extent of an offender’s problems and deficit areas. The weighted items of the needs assessment provided “a reasonably accurate relationship between the time needed for service delivery and overall need scores.” See Baird et al. (1979) at note 25, p. 14.

<sup>38</sup> Critics also see the risk-only tools as unhelpful for reassessments. Because the risk assessment component of these tools is often comprised of predominantly static risk items (e.g., criminal history), use of the same risk-only assessment after the offender has successfully completed a treatment intervention is unlikely to produce recidivism risk results much different from the offender’s original assessment. The CAIS addresses this issue by using a different risk assessment instrument for reassessment. C. Baird, personal communication, July 24, 2012. For the two different versions of the risk assessment instruments, see National Council on Crime and Delinquency. (2010). *CAIS Correctional Assessment and Intervention System: System manual*. Madison, WI: Author.

<sup>39</sup> See Skeem & Monahan commenting on the value of composite versus separate risk and needs instruments, assuming all have been validated:

If the ultimate purpose is to manage or reduce an individual’s risk, then value may be added by choosing an instrument that includes treatment-relevant risk factors. (Although an integrated instrument would be most parsimonious, we can easily envision a two-stage process in which a risk assessment step was followed by an independent risk management step.) This choice is appropriate for ongoing decisions in which the risk estimate can be modified to reflect ebbs and flows in an individual’s risk over time. Beyond focusing risk reduction efforts, these instruments could provide incentive for changing behavior (a parole board cannot advise an inmate to undo his past commission of an assault but can advise him to develop employment skills).” (p. 41)

Skeem, J. L., & Monahan, J. (2011). Current directions in violence risk assessment. *Current Directions in Psychological Science*, 20(1), 38-42. doi: 10.1177/0963721410397271

<sup>40</sup> Information on each item’s relationship to recidivism was not found in the studies reviewed for the COMPAS. The COMPAS *Practitioner’s Guide* provides information on the relationship between its risk scales and recidivism. See pp. 17-19 in Northpointe Institute for Public Management. (2013, January). *Practitioner’s guide to COMPAS*. Traverse City, MI: Author. Retrieved from

[http://www.northpointeinc.com/files/technical\\_documents/FieldGuide2\\_012813.pdf](http://www.northpointeinc.com/files/technical_documents/FieldGuide2_012813.pdf)

---

<sup>41</sup> Caudy et al. (2013) at note 34, p. 465: “In order for risk reduction strategies to be effective, needs assessments must be validated and linked to specific evidence-based interventions.”

<sup>42</sup> Vincent et al. (2012, November) at note 10, p. 34. See, also, Andrews & Bonta (2006) at note 2, p. 48.

<sup>43</sup> See, for example, Andrews, D.A., Bonta, J. L., & Wormith, J. S. (2009). *Level of Service/Case Management Inventory (LS/CMI) supplement: A gender-informed risk/need/responsivity assessment*. Toronto, ON: Multi-Health Systems, Inc.

<sup>44</sup> Bonta & Andrews (2007) at note 16. See, also, Crime and Justice Institute. (2004). *Implementing evidence-based principles in community corrections: The principles of effective intervention*. Boston: Author. Retrieved from <http://static.nicic.gov/Library/019342.pdf>

<sup>45</sup> For the LS/CMI, strengths are not included in the quantitative risk and need score. See Andrews et al. (2004) at note 9, p. 4. For information on the OST, see Arizona Adult Probation Services Division. (2009, July update). *OST Scoring Guide*. AZ: Authors.

<sup>46</sup> Because a risk-only instrument typically includes many static risk items, it is more quickly and easily scored. A probation department can screen an offender for risk right away and then conduct a needs assessment at a later date to inform case planning decisions. This approach also allows a jurisdiction to use the risk-only assessment to triage assessment administration resources. For example, if an offender is determined to be low-risk and therefore not an appropriate target for intensive risk-reduction treatment services, probation may determine that a full needs assessment is unnecessary whereas offenders determined to be moderate- or high-risk would be given a needs assessment. Composite risk and needs assessment instruments address this issue by providing or recommending a separate “quick screen” tool. See, for example, the ORAS Community Supervision Screening Tool, pp. 29-31 in Latessa, E. J., Smith, P., Lemke, R., Makarios, M., & Lowenkamp, C. T. (2009, July). *The creation and validation of the Ohio Risk Assessment System: Final report*. Cincinnati: Center for Criminal Justice Research, University of Cincinnati School of Criminal Justice. Retrieved from [http://www.ocjs.ohio.gov/ORAS\\_FinalReport.pdf](http://www.ocjs.ohio.gov/ORAS_FinalReport.pdf). A brief risk-only assessment may also be used “for expedited or early disposition cases to provide additional information to the court that otherwise would not be available because the person did not go through the presentence investigation process.” See p. 1 in Arizona Adult Probation Services Division (2009, July update). *MOST Scoring Guide*. AZ: Authors. Vincent et al. (2012, November) at note 10, pp. 58-59 caution that such screening tools should be used when risk is the only question; they should not be used to guide treatment planning.

<sup>47</sup> Some RNA instruments (e.g., LSI-R, LS/CMI, COMPAS) also provide information on another form of reliability referred to as *internal consistency*. Internal consistency reliability provides an indication of the extent to which all the items in a scale measure the same single underlying concept or dimension. The test commonly used to measure internal consistency is called Cronbach’s alpha. Because RNA tools are deliberately designed to measure multiple multifaceted factors related to recidivism rather than a single construct, test developers generally focus more on the tool’s predictive accuracy than on its internal consistency.

<sup>48</sup> See p. 286 in Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284-290.

<sup>49</sup> Other statistical techniques also capture inter-rater reliability, but are less commonly used in existing RNA research. The OST and WRN profiles report studies using percent agreement between raters.

---

<sup>50</sup> For more information about training and quality assurance, see Section 4 “What Practices Support Sound Implementation of Risk and Needs Assessment Instruments?”

<sup>51</sup> Examples of other forms of validity not covered in this Primer are **content validity**, or the degree to which the RNA tool measures all of the information that is conceptually relevant to a complete understanding of recidivism risk; **face validity**, or the degree to which the instrument makes intuitive sense to probation officers and other stakeholders (which can be important in motivating staff to actually use the tool); and **concurrent validity**, or the degree to which a new RNA tool reflects the same constructs measured by an existing or “gold standard” RNA tool.

<sup>52</sup> Grove, W., & Meehl, P. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, 2, 293-323 at 306.

<sup>53</sup> For more on this issue, see “*Use of administrative overrides*” in Section 4 “What Practices Support Sound Implementation of Risk and Needs Assessment Instruments?”

<sup>54</sup> See Gottfredson & Moriarty (2006) at note 12; Grove & Meehl (1996) at note 52; Latessa & Lovins (2010) at note 33; and Skeem & Monahan (2011) at note 39.

<sup>55</sup> Vincent (November 6, 2012) at note 10, pp. 81-82.

<sup>56</sup> See note 55.

<sup>57</sup> Researchers may report other statistics, such as Pearson’s chi-square ( $\chi^2$ ) test to determine how much separation an assessment tool achieves between risk level classifications or contingency tables with a Relative Improvement Over Chance (RIOC) value to quantify how much improvement the tool introduces over chance. However, these statistical techniques are less commonly reported.

<sup>58</sup> When creating or revalidating an RNA tool, researchers will often examine whether each item is related to recidivism, how each item in the assessment is weighted before the item scores are summed to create a raw recidivism risk score, and at which points in the continuum of raw risk scores could cutoffs be introduced to define the low, moderate, and high risk level classifications. For example, responses on each item in the original Wisconsin risk assessment determined the item’s score. After the item scores are summed to create a raw risk score, cutoff values of 8 and 15 were used to create the low, moderate, and high risk classification groups. See Baird et al. (1979) at note 25, p. 11. See, also, Baird (2009) at note 35, pp. 6-7, discussing the importance of examining recidivism rates by risk level in evaluating a risk assessment system.

<sup>59</sup> Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology*, 63, 737-748.

<sup>60</sup> See, for example, (1) Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, 62, 783-792. (2) Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, 1, 1-26. (3) Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen’s d, and r. *Law and Human Behavior*, 29, 615-620.

<sup>61</sup> The guidelines are based on the following works of Rice and Harris: Rice & Harris (1995) at note 59 and Rice & Harris (2005) at note 60. See, also, Hanson, R. K. (January, 2000). Risk assessment. Beaverton, OR: Association for the Treatment of Sexual Abusers. Retrieved from [http://www.cj-resources.com/CJ\\_Corrections\\_pdfs/InfoPac%20Risk%20assessment%20booklet%20-%20Hanson%202000.pdf](http://www.cj-resources.com/CJ_Corrections_pdfs/InfoPac%20Risk%20assessment%20booklet%20-%20Hanson%202000.pdf)

<sup>62</sup> Gendreau, P., Little, T., & Goggin, C. E. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology*, 34, 575-607. doi: 10.1111/j.1745-9125.1996.tb01220.x

---

<sup>63</sup> For example, as suggested by Hanson, even small effect sizes “may have considerable consequences in some contexts” (p. 176). Hanson, R. K. (2009). The psychological assessment of risk for crime and violence. *Canadian Psychology*, 50, 172-182.

<sup>64</sup> Hanson (2009) at note 63.

<sup>65</sup> See Rice & Harris (2005), at note 60, p. 619 and Hanson (2009) at note 63, p. 176.

<sup>66</sup> Researchers often examine subgroupings by particular offender demographic or descriptive characteristics like gender or race but may also examine differential validity by type of offense committed (e.g., among felony property offenders, felony drug offenders). Risk of violent crime reoffending and sex crime reoffending are often of particular interest to leaders and policymakers in the criminal justice system, but general risk assessment instruments typically are not developed and validated to address these specific forms of recidivism. Instead, specialized assessment tools have been developed specifically for estimating the likelihood that an offender will commit another violent crime or sex crime.

<sup>67</sup> Van Voorhis, P., Salisbury, E. J., Wright, E. M., & Bauman, A. (2008). *Achieving accurate pictures of risk and identifying gender responsive needs: Two new assessments for women offenders*. Washington, DC: National Institute of Corrections. Retrieved from

<http://www.uc.edu/content/dam/uc/womenoffenders/docs/NIC%20Summary%20Report.pdf>

See also Van Voorhis, P., Wright, E. M., Salisbury, E., & Bauman, A. (2010). Women’s risk factors and their contributions to existing risk/needs assessment: The current status of a gender-responsive supplement. *Criminal Justice and Behavior*, 37, 261-288. doi: 10.1177/0093854809357442

<sup>68</sup> Examples of criminogenic needs for women are parental stress, family support, anger, depression and other symptoms of mental illness, unsafe housing, educational assets, self-esteem, and self-efficacy. See Van Voorhis et al. (2008) at note 67, p. 14.

<sup>69</sup> For the LS/CMI, see Andrews et al. (2009) at note 43. For the COMPAS, see Brennan, T., Breitenbach, M., & Dieterich, W. (2008). *A need/risk explanatory classification of female prisoners incorporating gender-neutral and gender-responsive factors*. Traverse City, MI: Northpointe Institute for Public Management. Retrieved from

[http://www.northpointeinc.com/files/research\\_documents/A\\_Need-Risk\\_Explanatory\\_Classification\\_of\\_Females.pdf](http://www.northpointeinc.com/files/research_documents/A_Need-Risk_Explanatory_Classification_of_Females.pdf)

<sup>70</sup> A good tool often incorporates information from multiple methods of data collection, as this often results in gains in predictive validity. See Bonta (2002) at note 31.

<sup>71</sup> Northpointe Institute for Public Management (2013, January) at note 40, pp. 44-45.

<sup>72</sup> COMPAS also includes a Random Responding scale to identify offenders who may be randomly answering the questionnaire. See note 71.

<sup>73</sup> When choosing among existing RNA tools, some researchers recommend selecting a tool that has been evaluated by independent researchers in at least two separate studies. See Vincent et al. (November 6, 2012) at note 10.

<sup>74</sup> See p. 502 in Lowenkamp, C. T., & Latessa, E. J. (2005). Increasing the effectiveness of correctional programming through the risk principle: Identifying offenders for residential placement. *Criminology & Public Policy*, 4, 263-290. doi: 10.1111/j.1745-9133.2005.00021.x

<sup>75</sup> Lowenkamp, C. T., & Latessa, E. J. (2004). Understanding the risk principle: How and why correctional interventions can harm low-risk offenders. In National Institute of Corrections (Series Ed.), *Topics in Community Corrections: Assessment Issues for Managers* (pp. 3-8). Washington, DC: National Institute of Corrections.

<sup>76</sup> Johnson & Hardyman (2004) at note 26.

- 
- <sup>77</sup> See p. 479 in Ferguson, J. L. (2002). Putting the "What Works" research into practice: An organizational perspective. *Criminal Justice and Behavior*, 29, 472-492. Retrieved from <http://cjb.sagepub.com/content/29/4/472>.
- <sup>78</sup> White, T. F. (2004). Implementing an offender risk and needs assessment: An organizational change process. In National Institute of Corrections (Series Ed.), *Topics in Community Corrections: Assessment Issues for Managers* (pp. 42-48). Washington, DC: National Institute of Corrections.
- <sup>79</sup> Vincent et al. (November 6, 2012) at note 10, pp. 81-82.
- <sup>80</sup> That is, any RNA tool that generates a score for categorization purposes or which reports the probability of recidivism as a ratio or percentage likelihood. See note 79.
- <sup>81</sup> See p. 3 in Assessments.com. (2009). *The STRONG: Static Risk and Offender Needs Guide*. Bountiful, UT: Authors. Retrieved from [http://www.assessments.com/assessments\\_documentation/STRONG%20Fact%20Sheet.pdf](http://www.assessments.com/assessments_documentation/STRONG%20Fact%20Sheet.pdf)
- <sup>82</sup> B. Lovins, personal communication, February 16, 2012. R. Barnoski, personal communication, April 24, 2012.
- <sup>83</sup> Note that the independent researchers conducting this revalidation study also strongly recommended removal of an item that the original instrument developers acknowledged was not associated with recidivism but included in the original risk assessment instrument solely for policy reasons. Eisenberg, M., Bryl, J., and Fabelo, T. (July, 2009). *Validation of the Wisconsin Department of Corrections Risk Assessment Instrument*. New York: Council of State Governments Justice Center. Retrieved from <http://csgjusticecenter.org/wp-content/uploads/2012/12/WIRiskValidationFinalJuly2009.pdf>
- Baird et al. (1979) at note 25 and C. Baird, personal communication, July 24, 2012.
- <sup>84</sup> Clear, T. R., & Gallagher, K. W. (1985). Probation and parole supervision: A review of current classification practices. *Crime & Delinquency*, 31, 423-443. doi: 10.1177/001128785031003007
- <sup>85</sup> For example, one study confirmed the predictive validity of the LSI-R, but only when the assessment was scored by staff formally trained on how to properly administer the assessment. The relationship between LSI-R results and recidivism disappeared when untrained staff administered the tool. See Flores, A. W., Lowenkamp, C. T., Holsinger, A. M., & Latessa, E. J. (2006). Predicting outcome with the Level of Service Inventory-Revised: The importance of implementation integrity. *Journal of Criminal Justice*, 34, 523-529. doi: 10.1016/j.jcrimjus.2006.09.007
- <sup>86</sup> See *Guiding Principle 4: Stakeholder Training* (pp. 21-22) in Casey et al. (2011) at note 4.
- <sup>87</sup> Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida. Retrieved from <http://ctndisseminationalibrary.org/PDF/nirnmonograph.pdf>
- <sup>88</sup> Joyce, B., & Showers, B. (2002). *Student achievement through staff development* (3<sup>rd</sup> ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
- <sup>89</sup> Vincent et al. (2012, November) at note 10, p. 86.
- <sup>90</sup> Fixsen et al. (2005) at note 87.
- <sup>91</sup> See the STRONG instrument profile in the Appendix.
- <sup>92</sup> Refer to the *Override Policy* sections for each RNA tool featured in the Appendix.
- <sup>93</sup> B. Lovins, personal communication, December 7, 2012.
- <sup>94</sup> Grove & Meehl (1996) at note 52. See, also, Gottfredson & Moriarty (2006) at note 26.
- <sup>95</sup> See, for example, Arkes, H. R., Dawes, R. M., & Christensen, C. (1986). Factors influencing the use of a decision rule in a probabilistic task. *Organizational Behavior and Human Decision*
-

---

Processes, 37, 93-110. Also, Goldberg, L. R. (1968). Simple models or simple processes? Some research on clinical judgment. *American Psychologist*, 23, 483-496.

<sup>96</sup> Wormith, J. S., Hogg, S., & Guzzo, L. (2012). The predictive validity of a general risk/needs assessment inventory on sexual offender recidivism and an exploration of the professional override. *Criminal Justice and Behavior*, 39, 1511-1538. doi: 10.1177/0093854812455741

<sup>97</sup> See, for example, Center for Sentencing Initiatives (December, 2013). *Use of risk and needs assessment information at sentencing: Napa County, California*. Williamsburg, VA: National Center for State Courts. Retrieved from

[http://www.ncsc.org/sitecore/content/microsites/csi/home/Topics/~/\\_media/Microsites/Files/CSI/RNA%20Brief%20-%20Napa%20County%20CA%20ocsi.ashx](http://www.ncsc.org/sitecore/content/microsites/csi/home/Topics/~/_media/Microsites/Files/CSI/RNA%20Brief%20-%20Napa%20County%20CA%20ocsi.ashx)

Center for Sentencing Initiatives (December, 2013). *Use of risk and needs assessment information at sentencing: Grant County, Indiana*. Williamsburg, VA: National Center for State Courts. Retrieved from

[http://www.ncsc.org/sitecore/content/microsites/csi/home/Topics/~/\\_media/Microsites/Files/CSI/RNA%20Brief%20-%20Grant%20County%20IN%20ocsi.ashx](http://www.ncsc.org/sitecore/content/microsites/csi/home/Topics/~/_media/Microsites/Files/CSI/RNA%20Brief%20-%20Grant%20County%20IN%20ocsi.ashx)

<sup>98</sup> See, for example, Center for Sentencing Initiatives (December, 2013). *Use of risk and needs assessment information at sentencing: Mesa County, Colorado*. Williamsburg, VA: National Center for State Courts. Retrieved from

[http://www.ncsc.org/sitecore/content/microsites/csi/home/Topics/~/\\_media/Microsites/Files/CSI/RNA%20Brief%20-%20Mesa%20County%20CO%20ocsi.ashx](http://www.ncsc.org/sitecore/content/microsites/csi/home/Topics/~/_media/Microsites/Files/CSI/RNA%20Brief%20-%20Mesa%20County%20CO%20ocsi.ashx)

Center for Sentencing Initiatives (December, 2013). *Use of risk and needs assessment information at sentencing: 7<sup>th</sup> Judicial District, Idaho*. Williamsburg, VA: National Center for State Courts. Retrieved from

[http://www.ncsc.org/sitecore/content/microsites/csi/home/Topics/~/\\_media/Microsites/Files/CSI/RNA%20Brief%20-%207th%20Judicial%20District%20ID%20ocsi.ashx](http://www.ncsc.org/sitecore/content/microsites/csi/home/Topics/~/_media/Microsites/Files/CSI/RNA%20Brief%20-%207th%20Judicial%20District%20ID%20ocsi.ashx)

<sup>99</sup> Kreamer, S. (2004). Quality assurance and training in offender assessment. In National Institute of Corrections (Series Ed.), *Topics in Community Corrections: Assessment Issues for Managers* (pp. 13-19). Washington, DC: National Institute of Corrections.

<sup>100</sup> Vincent et al. (2012, November) at note 10, p. 84.

<sup>101</sup> See the ORAS instrument profile in the Appendix.

<sup>102</sup> Andrews et al. (2004) at note 9, p. 6.

<sup>103</sup> Ferguson (2002) at note 77.

<sup>104</sup> Vincent et al. (2012, November) at note 10, p. 57.

<sup>105</sup> Vincent et al. (2012, November) at note 10. See, also, Del Pra, Z. (2004). In search of a risk instrument. In National Institute of Corrections (Series Ed.), *Topics in Community Corrections: Assessment Issues for Managers* (pp. 9-12). Washington, DC: National Institute of Corrections.

<sup>106</sup> Ferguson (2002) at note 77.

---

## APPENDIX

---

### *Risk and Needs Assessment Instrument Profiles*

This Appendix reviews six risk and needs assessment (RNA) tools. As explained in the Primer, each profile begins with a glossary of definitions for common terms used in the creation of the RNA tools. The terms and their definitions vary somewhat across the tools. The profiles also present information on the following general categories: (a) history and current use, (b) development, (c) content, (d) instrument reliability and validity, and (e) practical considerations. The profiles are based on a review of the literature and interviews with at least one individual involved in the development of each instrument. The instrument developers also had an opportunity to respond to a discussion guide prepared for each instrument that was revised following each interview as well as the final draft versions of the profiles.

Readers are encouraged to read the Primer to gain a broader context regarding the purpose and appropriate use of RNA tools and a better understanding of some of the terms (e.g., reliability and validity) used in the profiles.

<i>RNA Instrument Profile</i>	<i>Page</i>
Correctional Assessment and Intervention System (CAIS)	A-2
Correctional Offender Management Profile for Alternative Sanctions (COMPAS)	A-19
Level of Service Assessments (LSI-R and LS/CMI)	A-30
Offender Screening Tool (OST)	A-44
Ohio Risk Assessment System (ORAS)	A-52
Static Risk and Offender Needs Guide (STRONG)	A-59

# Correctional Assessment And Intervention System (CAIS)\*

---

### CAIS GLOSSARY OF TERMS

---

<b><i>Risk</i></b>	Risk refers to the aggregate likelihood that an offender classified into a particular risk group will commit subsequent criminal behavior. <sup>1</sup>
<b><i>Static risk</i></b>	Christopher Baird, a National Council on Crime and Delinquency (NCCD) CAIS author, recognizes the use of static (i.e., not changeable) and dynamic (i.e., changeable) factors by the field but did not use those distinctions in developing the instrument: “Any factor (other than those that should not be included for ethical reasons) that adds to the instrument’s ability to optimally separate risk groups should be included in a risk tool. It does not matter if a factor is static or dynamic.” <sup>2</sup>
<b><i>Dynamic risk</i></b>	See above. Baird describes the CAIS system in its entirety as dynamic. “At reclassification, the emphasis shifts from prior criminal history items to measures that reflect adjustment during supervision,” allowing “clients to move between supervision levels based on their performance.” <sup>3</sup>
<b><i>Needs</i></b>	Needs refer to “problems and deficit areas” most commonly evidenced in probationers and parolees. <sup>4</sup> According to Baird, a particular need is not criminogenic (i.e., causing criminal behavior) in and of itself; rather a need can only be deemed criminogenic for an individual offender. <sup>5</sup>
<b><i>Responsivity</i></b>	Term not used explicitly in reports on the creation of CAIS.
<b><i>Protective factors</i></b>	Term not used in reports on the creation of CAIS. Baird contends that protective factors can be important to case planning and management but is critical of the manner in which these factors have been assessed and used by the field. <sup>6</sup>
<b><i>Strengths</i></b>	CAIS considers strengths and needs in developing supervision strategies. Potential strengths are areas rated by the interviewer as having no or only minor significance in generating criminal behavior. <sup>7</sup>
<b><i>Recidivism</i></b>	The CAIS manual defines recidivism as “the likelihood that an offender will experience a subsequent felony conviction or be revoked into an institutional setting in the next 24 months.” <sup>8</sup>

---

\*The CAIS is based on components of the National Institute of Corrections’ Model Probation and Parole Management Program, including the Wisconsin Risk and Needs (WRN) assessment instruments and the Client Management Classification (CMC) planning guide. Accordingly, these also are discussed in the profile.

## Appendix: RNA Instrument Profile for the CAIS

---

### HISTORY & CURRENT USE.

---

**Creation.** The Correctional Assessment and Intervention System (CAIS) evolved from efforts in Wisconsin, beginning in 1975 at the direction of the state legislature, to develop a case classification system for probationers and parolees that would improve the effectiveness of service delivery.<sup>9</sup> Though the Wisconsin effort began with funding from the Law Enforcement Assistance Administration (LEAA), it required four years and substantial additional resources from the Wisconsin Division of Corrections, Bureau of Community Corrections to design, implement and evaluate.<sup>10</sup> The classification system that emerged from this effort, commonly referred to as the Wisconsin Risk and Needs (WRN) assessment, has separate risk assessment and needs assessment components, each developed independent of the other using different methodologies.<sup>11</sup> The risk and needs scores were used principally to determine an appropriate level of supervision for an offender but did not address case planning and supervision. To address this gap, the Client Management Classification (CMC) system was developed.<sup>12</sup> The CMC uses information about offender needs, as well as other factors thought to distinguish different types of offenders, to classify an offender into one of four supervision categories.

The CAIS combines updated versions of the Wisconsin risk, needs, and supervision strategy assessments into a single, automated system to assist case managers with the effective and efficient supervision of offenders.<sup>13</sup> CAIS provides this information through a web-based data system accessible

via internet browser. In addition to providing individual offender assessment reports, CAIS also has the capability to produce aggregate, managerial reports to help identify service gaps and target resources.<sup>14</sup> Much of the information and research available is on earlier versions of the various CAIS components. Thus this profile reviews the development and application of the WRN and CMC as the precursors to the CAIS.

**Current use.** Numerous correctional agencies outside of Wisconsin adopted the WRN (or a slight variation of the instrument) and the CMC after the instruments became part of the National Institute of Corrections (NIC) Model Probation and Parole Project in the 1980s.<sup>15</sup> A survey of 288 state and local probation and parole agencies by the University of Cincinnati in 1998-1999 reported that the most widely used instrument is the CMC system (36%), including both the WRN instruments; another 26.3% reported using the WRN assessment but not the CMC; 2% reported using only the Wisconsin risk assessment, and less than 1% reported using the Wisconsin needs assessment alone.<sup>16</sup> In addition, inspection of the instruments falling in the “other” category also revealed that some of these instruments were versions of the WRN assessments.

The National Survey of Criminal Justice Treatment Practices, a survey of prisons, jails, and community correctional agencies begun in 2002, identified the WRN as the second most frequently used assessment instrument by these agencies, though the percentage was only 12.7% because nearly two-thirds of the facilities reported not using

## Appendix: RNA Instrument Profile for the CAIS

---

any instrument.<sup>17</sup> It is not known how well these figures reflect current use of the instruments.

CAIS, more recent in its development, is currently used by ten agencies, including county probation departments, a county jail with an associated reentry program, a county reentry program and a non-governmental community-based reentry program.<sup>18</sup>

### DEVELOPMENT.

---

***Instrument purpose.*** CAIS “is a supervision strategy model that weaves together a risk assessment and a needs assessment.”<sup>19</sup> CAIS identifies the underlying motivation for an offender’s criminal behavior to assist in developing the offender’s case plan.

According to its developers, its purpose is to assist case managers with supervising offenders effectively and efficiently with the goals of aiding institutional adjustment, reducing recidivism, and helping offenders live productively in the community.<sup>20</sup>

***Approach to instrument development.***

CAIS is designed to accommodate a variety of risk assessment instruments, but the default instrument is a modified version of the Wisconsin Department of Corrections risk assessment instrument (sometimes referred to as the DOC-502 risk scale).<sup>21</sup> The Wisconsin risk assessment was developed using a criterion variable that combined the number of occurrences of absconsions, rules violations, arrests, misdemeanor convictions, felony convictions, and convictions for assaultive offenses.<sup>22</sup> Utilizing a retrospective design, information was collected on approximately 250 randomly selected closed or revoked cases. A working committee of

probation officers, supervisors and research staff identified 22 items they associated with offender recidivism based on professional judgment and consensus opinion.

Researchers then applied linear regression techniques to refine this pool of items and eliminate items that failed to demonstrate a statistically significant relationship with recidivism. Seven items were retained as a result of this process. To enhance predictive validity, researchers added three items that were not identified by the regression analysis but nonetheless had a strong relationship with the outcome measure (examining item significant differences and simple correlation coefficients) and discriminated among high, moderate, and low risk offenders.<sup>23</sup> The final scale consisted of these ten items, each weighted based on its correlation with criminal behavior.

At the explicit request of the Wisconsin Department of Corrections, the test developers added an eleventh item, history of assaultive offense, to the instrument. The purpose of this item was to ensure that offenders “who had committed an assaultive offense within the last five years are placed under maximum supervision for (at least) the first six months of probation or parole.”<sup>24</sup> The item added 15 points to an offender’s risk assessment score, the minimum score needed to be placed under maximum supervision. At reevaluation, supervision levels were based solely on risk and needs scores; the additional points were not added to the offender’s reevaluation score.<sup>25</sup> The additional assaultive item was never considered to be part of the ten-item actuarial risk scale because it was never shown to be related to the risk of recidivism.<sup>26</sup> However, the item was

## Appendix: RNA Instrument Profile for the CAIS

---

included in some subsequent use of the risk scale by others despite its lack of predictability in the development of the instrument.

After construction, the risk scale was initially tested on a sample of 4,231 Wisconsin offenders. The results indicated that initial risk scores were related to subsequent revocations: Approximately 2% of low risk offenders, 9% of moderate risk offenders, and 26% of high risk offenders were revoked.<sup>27</sup>

The needs component of the WRN assessment, and subsequently the CAIS, was designed to assess the extent of an offender's problems and deficit areas to better estimate the amount of supervision time the case would require.<sup>28</sup> The Wisconsin project also sought to standardize the needs information collected across probation officers.<sup>29</sup>

To develop the needs tool, probation and parole officers and researchers identified an extensive list of possible client needs and, using the list, surveyed incoming clients over an eight-month period in Madison. A set of eleven areas of needs emerged from this process: 1) academic/vocational skills, 2) employment problems, 3) financial management, 4) marital/family relationships, 5) companions, 6) emotional stability, 7) alcohol use, 8) other drug use, 9) mental ability, 10) health, and 11) sexual behavior.<sup>30</sup> Together, these areas were "thought to encompass the wide range of problems that are most commonly evidenced in probationers and parolees."<sup>31</sup>

Each of the eleven items and a twelfth item assessing the probation officer's impression of the offender's needs is weighted based on

supervision time to address the need. Initially based on the professional judgment of the probation and parole agents, the weights were subsequently empirically verified on a sample of 482 offenders as presenting "a reasonably accurate relationship between the time needed for service delivery and overall need scores."<sup>32</sup>

In Wisconsin, agencies used the highest score of either the risk or needs scale to determine the level of supervision.<sup>33</sup> When other states began using the instrument, this practice varied with some states relying more on one or the other instrument—usually the risk assessment.<sup>34</sup> Eventually, most users settled on the risk assessment for determining level of supervision, as is the approach taken with the CAIS.<sup>35</sup>

Once the level of supervision is known, probation officers turned to CMC to develop a case plan and supervision strategy for an offender.<sup>36</sup> The CMC was developed by two clinical psychologists, a line officer, and research staff.<sup>37</sup> The development team began by identifying items with a potential for differentiating among basic offender types. They used the items to create an instrument based on forced-choice ratings, i.e., each item has several possible choices, and the interviewer selects the choice that best describes the offender. To increase the reliability of ratings, the team developed a 45-minute semi-structured interview with scripted questions and a companion scoring guide. The development process eventually yielded 45 offender attitude questions, 11 objective background and offense history items, 8 interview behavior items, and 7 interviewer impression items.

## Appendix: RNA Instrument Profile for the CAIS

---

The development team identified four supervision strategies based on their extensive experience working with offenders.<sup>38</sup> The team assessed a sample of offenders and, based on the assessment, subjectively placed each offender into one of the supervision strategies. The CMC items were then tested to see how strongly each influenced the professionals' decisions.<sup>39</sup> Weights were assigned to each item based on its ability to discriminate among the supervision strategy groups and its interrater reliability score.<sup>40</sup>

The test developers tracked 250 offenders in both the construction and cross-validation samples for 12 months to determine if offender behaviors were consistent with the expected problems and needs associated with the supervision strategy to which they were assigned.<sup>41</sup> The CMC system was modified to improve its reliability and validity based on the resulting data.<sup>42</sup> Using the data and their knowledge of supervision strategies, the test developers created supervision guidelines for the offenders in each strategy group. The guidelines provided information on "offender goals, officer/offender relationships, appropriate auxiliary services and programs, and supervision techniques."<sup>43</sup>

In explaining the development of the CAIS, the National Council on Crime and Delinquency (NCCD) noted that probation and parole agencies had become discontented with the CMC because it was not automated.<sup>44</sup> As a result, NCCD embarked on a two-year process to update and automate the CMC, resulting in the CAIS. CAIS incorporates the Wisconsin risk scale or other validated actuarial risk

assessment, needs items and CMC items.<sup>45</sup> Thus information collected in one interview provided probation officers with an offender's risk level, suggested supervision strategy, and principal service needs. In creating the CAIS, a few items were added (e.g., "What was your behavior like?" was added as a follow-up to "How would you describe yourself as a child?") or revised (e.g., "How much socializing do you do with women (men)?" revised to "Can you tell me about your relationships with women/men?"). In addition, several items (e.g., "Do you have any children?" and "How do you feel about being a mom?") were added for assessment of female offenders. NCCD reports that CAIS developers relied on an expert in gender issues to help develop gender-specific supervision strategies that focus on programs shown to be effective with female offenders.<sup>46</sup> As a result, the supervision and case planning recommendations may be somewhat different than what came out of the original system.<sup>47</sup>

### CONTENT.

---

**Structure.** CAIS generates a report that consists of two sections: Primary Case Planning Approach and the Specific Client Profile.<sup>48</sup> The Primary Case Planning Approach section has five sub-sections: (1) classification (providing scores for each supervision strategy and identifying the primary strategy to follow); (2) general issues facing offenders in the selected strategy; (3) goals of supervision; (4) common needs/referrals for offenders in the supervision strategy; (5) caseworker/offender relationship (providing guidance for working

## Appendix: RNA Instrument Profile for the CAIS

with offenders in the particular supervision strategy); and (6) techniques of supervision (i.e., those that are particularly applicable for the specific supervision strategy).<sup>49</sup>

The Specific Client Profile consists of three sections: (1) risk level; (2) principal service needs; and (3) special concerns.

**Items and domains.** Table 1 summarizes the number of items for each of the four major sections of the CAIS instrument as administered to female and male offenders.

**Table 1. CAIS Sections**

CAIS Sections	# of Items	
	Women	Men
1. General Information		
• Offense Patterns	8	8
• School Adjustment	5	5
• Vocational and Residential Adjustment	7	7
• Family Information	19	17
• Interpersonal Relations	7	7
• Feelings	6	6
• Plans and Problems	5	5
2. Objective History	11	11
3. Behavioral Observations	8	8
4. Interviewer Impressions	12	8
Total Number of Items	88	82

The 11 risk items are embedded within the “General Information” and “Objective History” sections. They are the same for female and male offenders. As noted earlier, however, jurisdictions can opt to replace the default CAIS risk assessment with their own

validated risk instrument if they prefer. The risk items for the CAIS and the original Wisconsin risk instrument are compared in Table 2.<sup>50</sup>

**Table 2. CAIS and Wisconsin Risk Items<sup>51</sup>**

CAIS Risk Items	WRN Risk Items
1. Employment	1. % of time employed in last 12 months
2. Address changes in the last year	2. Address changes in last 12 months
3. Offender’s pattern of associates	
4. Age at first arrest	3. Age at first conviction
5. # of prior offenses	4. # of prior felony convictions
6. Ever convicted for theft, burglary, auto theft, robbery	5. Convictions for burglary, theft, auto theft, robbery, worthless checks or forgery
7. # of prior jail sentences	
8. # of prior periods of probation or parole supervision	6. # of prior periods of probation/parole supervision
9. Ever had probation or parole revoked	7. # of prior probation/ parole revocations
10. % of criminal behavior related to alcohol abuse	8. Alcohol usage problems
11. % of criminal behavior related to other drug use	9. Other drug usage problems
	10. Attitude

Unlike the original Wisconsin needs assessment instrument, the CAIS does not provide an overall need score; rather it identifies areas that should be addressed in the offender’s case plan.

---

## Appendix: RNA Instrument Profile for the CAIS

---

Following the CMC approach, CAIS classifies offenders into one of four supervision strategies. According to the CAIS manual, the classification is based on items from all sections of the CAIS.<sup>52</sup> The manual also explains that “scores for the strategy groups are the result of a complex set of research-based scoring rules.”<sup>53</sup> The four supervision groups are:<sup>54</sup>

- The Selective Intervention (SI) strategy, which includes different strategies for situational (SI-S) and treatment (SI-T) groups, is for offenders who generally have pro-social values, positive adjustment, positive achievements, and good social skills.
- The Casework/Control (CC) strategy is for offenders with a broad range of instability, a chaotic lifestyle, emotional instability, multi-drug abuse/addiction, and negative attitudes towards authority.
- The Environmental Structure (ES) strategy is for offenders who lack social and survival skills, have poor impulse control, are gullible and naïve, and show poor judgment.
- The Limit Setting (LS) strategy is for offenders with antisocial values, who prefer to succeed outside the rules/law, whose role models operate outside the rules/law, and are manipulative and exploitive.

**Reporting risk levels.** The CAIS groups offenders into three levels of risk: low, moderate and high. The CAIS provides initial ranges of scores for each risk level; however, the NCCD, which holds the copyright to the CAIS, reports that “as part of each CAIS implementation project, NCCD validates the risk instrument periodically and customizes

the instrument for each agency to ensure it optimally classifies cases.”<sup>55</sup> NCCD encourages agencies to collect reassessment data which provides information on the current status of a case to assist with validation.<sup>56</sup>

### INSTRUMENT RELIABILITY AND VALIDITY.

---

NCCD does not indicate whether the updated and gender-specific versions of the CAIS were evaluated independently. The studies cited in support of the CAIS, and presented in the following sections, are those based on the original CMC.

In addition, Baird describes the Wisconsin risk and needs scales as providing an approach to assessing offender risk and needs.<sup>57</sup> The intent was to provide templates that jurisdictions could customize for their particular populations based on their own validation studies. As a result, there are many versions of the risk and needs scales with minor variations, which should be taken into consideration when comparing the results of validation studies across jurisdictions.

**Populations studied.** In addition to the statewide Wisconsin construction and validation samples of probation-eligible male and female adult offenders, the Wisconsin risk and needs assessment instruments and the CMC have been implemented and studied in a variety of states and Canada.

**Predictive validity.** Gendreau and his colleagues reported a mean effect size of  $r=.27$  between the Wisconsin risk scale and measures of recidivism.<sup>58</sup> The meta-analysis

## Appendix: RNA Instrument Profile for the CAIS

---

was based on 14 effect sizes calculated from various studies. It is not known how many of the effect sizes were based on the instrument with the assaultive factor included versus excluded.<sup>59</sup> Bonta reported correlations of  $r=.22$  to  $r=.33$  between risk scores and recidivism across a 7-year period for probationers in Manitoba, Canada. The analyses defined recidivism as failure on probation for technical violations and new offenses and were based on over 14,000 offenders on probation between 1986 and 1991.<sup>60</sup> The report does not indicate whether the assaultive item was included on the scale.

More recently, Eisenberg and his colleagues examined the performance of the Wisconsin risk instrument for a sample of 42,853 Wisconsin offenders placed on community supervision in 2001-2002.<sup>61</sup> They found a correlation of  $r=.22$  between risk scores (excluding the assaultive factor) and the commitment of a new offense within three years of being placed on community supervision. Henderson and Miller examined a sample of 194 male, mostly misdemeanor offenders, released in 2000 from a Texas probation department. For the risk assessment with the assaultive item, they reported a correlation of  $r=.25$  (and an AUC of .63 for the receiver operating characteristic curve analysis) for arrest within five years of release from probation.<sup>62</sup> Latessa and his colleagues reviewed arrests for a new crime for 672 individuals on community supervision in Ohio in 2008. They found the correlation between the Wisconsin risk assessment and recidivism to be  $r=.21$ .<sup>63</sup> The researchers did not indicate whether the assaultive factor was included in the assessment, but Baird reports that the

results are actually based on the reassessment and needs instrument combined rather than the intake assessment.<sup>64</sup> The assaultive factor is not included in the reassessment instrument, and there are other differences between the two versions as well.

Several studies considered how accurately the Wisconsin risk assessment classified offenders into different risk levels as measured by subsequent recidivism. For example, revalidation studies for the Department of Corrections in Nevada and Wisconsin and for the probation departments in Orange County, California and Travis County, Texas all indicated that the recidivism rate for offenders increased with increasing classification levels of risk.<sup>65</sup> That is, offenders classified as low risk based on the Wisconsin risk scale recidivated less than offenders classified as medium risk, and both recidivated less than those classified as high risk. The Travis County revalidation included the assaultive factor (giving it a weight of 8 points) in its risk scale as did the Wisconsin revalidation (giving it a weight of 15 points). The Travis County report concluded the assaultive factor was predictive of recidivism, and the Wisconsin report concluded the factor did not adequately predict recidivism. The Nevada, Wisconsin, and Orange County reports all suggested revisions to the instrument to increase its ability to distinguish across risk levels. For example, the Orange County study indicated that a large percentage of offenders (54.8%) were classified as high risk. The study's authors suggested changing the weights for three items, eliminating one, adding a new item, and changing the cutoff scores for the classification levels. As a result

## Appendix: RNA Instrument Profile for the CAIS

---

of these changes, the percentage of offenders classified as high risk decreased to 34.5% while maintaining increasing levels of recidivism rates across the low, medium, and high classifications, and the AUC increased from .642 for the original instrument to .659 for the new instrument.<sup>66</sup>

A few studies have examined the relationship between the original Wisconsin needs assessment scale and recidivism and have found that some of the needs items are significantly related to recidivism.<sup>67</sup> However, the needs assessment scale was not specifically developed as a predictor of recidivism, and the CAIS does not report a separate needs score.

Researchers involved in the development of the CMC reported on an evaluation of the CMC in a summary article in 1986.<sup>68</sup> The evaluation followed 422 high-risk (as determined by the Wisconsin risk assessment) Milwaukee probationers randomly assigned to regular supervision, intensive supervision only, or intensive supervision as directed with CMC case planning. The study focused on three outcome measures: percentage revoked, percentage employed at termination and percentage earning income over \$400/month at termination. Although the results were in the predicted direction—the CMC with intensive supervision group performed better than the intensive supervision only group, and both performed better than the regular supervision group—only the comparison between the CMC with intensive supervision group and the regular supervision group was significant.

Researchers from the Texas Board of Pardons and Paroles followed 2,551 parolees, released during March and April 1985, for a year.<sup>69</sup> A little less than half (46%) of the parolees were supervised by parole officers trained on the CMC, and the remaining parolees served as a comparison group. All of the cases were classified as a poor, fair, or good risk based on a validated risk assessment. The CMC parolees had significantly fewer pre-revocation warrants than regular supervision parolees for the poor and fair risk groups when measured after 6- and 12-month periods. CMC parolees in the poor risk group also had significantly fewer returns to prison than non-CMC parolees. Thus CMC had the greatest effect on high risk offenders; no statistical difference was found for parolees in the good risk categories.

Researchers from the South Carolina Department of Probation, Parole and Pardon Services also found that CMC was related to outcomes for higher risk offenders—those convicted of a violent or sexual offense, who have served more than 90 days in prison, or who are under intensive supervision.<sup>70</sup> They followed two groups of offenders, matched on the basis of offense, risk score, and level of supervision, for a year during 1985-1986. One group of 200 offenders was supervised with CMC, and the other group of 219 offenders was not. The two groups differed significantly on measures of supervision failure, revocations for new offense, and revocations or unsatisfactory supervision terminations resulting in returns to prison.

CMC developers also report data from an unpublished study of 45,346 offenders in Florida placed in a community control program as an alternative to prison.<sup>71</sup>

## Appendix: RNA Instrument Profile for the CAIS

---

Approximately half of the offenders received CMC in addition to the supervision requirements for all offenders. Data for the first four years (1993 to 1997) of the program indicated that the offenders supervised with CMC had significantly lower revocation rates.

Harris and her colleagues, however, questioned the use of revocation rates as the primary indicator of success for the CMC. They suggested that officers trained on CMC techniques may be less likely to revoke offenders. They assessed the effectiveness of the CMC using three different outcome measures: write-ups for technical violations, revocations, and new arrests while under supervision. Of the 1,017 felony offenders entering probation for approximately a year beginning in March of 1991, 581 were supervised with CMC, and 436 served as the control group. CMC-supervised offenders differed significantly from offenders in the control group only on the outcome measure of revocations. In addition, the CMC group had a higher failure to comply with program conditions despite being less likely to experience revocation compared to the control group. However, an audit of the CMC-supervised cases indicated errors in implementation by probation officers, thus calling into question the extent to which CMC was implemented as intended. The authors called for more evaluations of CMC using multiple outcome measures to ensure successful revocation outcomes are due to changes in offenders' behaviors and not to officers' more tolerant supervision strategies regarding revocations for minor infractions.

**Reliability.** No information was found on the inter-rater reliability of the Wisconsin

risk instrument. Both the Wisconsin and Orange County validation studies recommended conducting inter-rater and intra-rater reliability testing to assure accurate scoring.<sup>72</sup>

The inter-rater reliability of the original needs scale was examined during its development. Probation officers listened to taped interviews with offenders and independently rated the needs of the offenders. The average rate of agreement for each of the eleven items ranged between 79% and 94% with an average overall rating of 87%.<sup>73</sup>

The report on the development of the CMC indicates that "different raters obtain the same client groups approximately 90% of the time," and agreed on individual items 70% of the time or higher, with a few exceptions.<sup>74</sup>

**Potential for bias: gender.** The revalidation of the risk instrument in Nevada; Wisconsin; Orange County, California; and Travis County, Texas all found that the instrument performed as expected for both males and females.<sup>75</sup> That is, recidivism increased across low, medium, and high categories of risk for males and females. However, as discussed under the "predictive validity" section, suggestions were made to revise the scale and cutoff scores for risk levels to improve the classification categories for all offenders.

According to the CAIS brochure, the assessment system includes "gender-specific system factors in the unique risk and needs areas of women as well as tailoring supervision strategies for women based on the most current research."<sup>76</sup> Studies comparing the recidivism rate of female

## Appendix: RNA Instrument Profile for the CAIS

---

offenders supervised based on their CAIS assessment versus those supervised without CAIS are not available in the general literature to date.

**RACE.** The revalidations of the risk instrument in Nevada; Wisconsin; and Orange County, California also found that the instrument performed as expected for Black, White, and Hispanic groups.<sup>77</sup> As with gender, suggestions were made to revise the scale and cutoff scores for risk levels to improve the classification categories for all offenders.

Studies comparing the recidivism rate of different race and ethnic groups supervised based on their CAIS assessment versus those supervised without CAIS are not available in the general literature to date.

**Independent validation.** Several of the studies cited in the validation section were conducted by independent researchers. In addition, NCCD has conducted or reported on several unpublished validation studies.

### **PRACTICAL CONSIDERATIONS.**

---

**Vendor and instrument cost.** The original Wisconsin risk and needs scales and CMC are in the public domain. CAIS is proprietary. The automated assessment and case management system is available for purchase from AutoMon and NCCD.<sup>78</sup> A subscription fee is assessed of users, the amount of which is determined on a sliding scale based on the size of the jurisdiction.<sup>79</sup> For more information, contact NCCD at [JAIS.CAIS@nccdglobal.org](mailto:JAIS.CAIS@nccdglobal.org) or AutoMon at [sales@automon.com](mailto:sales@automon.com).

**Menu of other services.** NCCD and AutoMon offer a wide array of services, training, and technical assistance to support CAIS implementation.

- **IT SERVICES.** CAIS is a web-based program available through an internet browser. The advantage of this approach is that there are no issues with infrastructure requirements and redesign of existing agency MIS systems.<sup>80</sup> AutoMon is a computer software firm that provides technology support and can customize the system to include additional assessment tools and specific reports.
- **TECHNICAL ASSISTANCE.** NCCD offers two technical assistance visits of two days each per year to CAIS clients.<sup>81</sup>
- **VALIDATION SERVICES.** NCCD will validate the risk assessment component of the CAIS for all client agencies as part of the package of services provided.<sup>82</sup> NCCD recommends conducting a revalidation study every 2-5 years, depending on the size of the jurisdiction (smaller jurisdictions may need a longer period of time to identify a large enough cohort of cases for a revalidation study). There is no added cost for this service.
- **USER TRAINING.** NCCD offers a training package that includes 24 hours of classroom work and additional follow-up practicum work.<sup>83</sup> The training is fee-based. An optional 3-day “train the trainers” course is also available and is recommended for those clients interested in developing an internally sustainable initial and refresher training program. In addition, web-based courses

## Appendix: RNA Instrument Profile for the CAIS

---

have been developed “to reduce training costs and provide greater flexibility to agencies to train new staff or provide refresher training when needed.”<sup>84</sup>

**User qualifications.** All users must take the mandatory training (see description, above) before using the CAIS.

**Administration time.** The CAIS manual reports that an assessment generally takes approximately 45 minutes to complete.<sup>85</sup>

**Modes of administration.** A semi-structured interview format is used to complete the CAIS. The CAIS manual encourages officers to follow-up on important or interesting information the offender presents during the interview.<sup>86</sup>

**Quality assurance.** When adopting any offender assessment tool, jurisdictions must be prepared to ensure appropriate implementation and proper maintenance over time. Quality assurance recommendations and guidelines for CAIS follow.

- **OVERRIDE POLICY.** The CAIS report provides an opportunity for the officer to override the risk level based on a state or local policy or at the officer’s discretion, provided a reason is given and a supervisor approves the override. The reasons for overrides vary across jurisdictions. Though some jurisdictions have made extensive use of the policy override (e.g., certain offenses automatically are placed in higher risk levels, as discussed in previous sections), discretionary overrides are less frequent. Baird reports that NCCD studies usually see overrides in the 5-7% range.<sup>87</sup>

- **FIDELITY.** CAIS offers a variety of aggregate data report options for officers and supervisors. Information regarding the implementation of the CAIS can be routinely obtained and reviewed on issues such as gender, risk levels, needs, ethnicity, worker, and unit.<sup>88</sup>
- **INSTRUMENT REVALIDATION.** Validations of the risk component are recommended every 2-5 years, depending on the size of the jurisdiction and available data.

### ENDNOTES

---

<sup>1</sup> “The goal of risk assessment is to classify offenders into different risk groups based on rates of subsequent criminal behavior” (C. Baird, personal communication, July 24, 2012). Various individuals have been involved in the development and evaluation of the different components of the CAIS; however, Christopher Baird of the National Council on Crime and Delinquency has written the most about the instrument’s development and use.

<sup>2</sup> C. Baird, personal communication, July 24, 2012.

<sup>3</sup> See p. 38 in Baird, C. (1981). Probation and parole classification: The Wisconsin model. *Corrections Today*, 43(3), 36-41. Information also provided by C. Baird, personal communication, July 24, 2012.

<sup>4</sup> See p. 12 in Baird, C. S., Heinz, R. C., & Bemus, B. J. (1979). *The Wisconsin case classification/staff deployment project: A two-year follow-up report*. Madison, WI: Wisconsin Division of Corrections.

<sup>5</sup> See pp. 8-9 in Baird, C. (2009). *A question of evidence: A critique of risk assessment models used in the justice system*. Oakland, CA: National Council on Crime and Delinquency.

<sup>6</sup> See Baird (2009) at endnote 5, pp. 9-10.

## Appendix: RNA Instrument Profile for the CAIS

---

<sup>7</sup> See pp. 2, 69 in National Council on Crime and Delinquency. (2010). *CAIS Correctional Assessment and Intervention System: System manual*. Madison, WI: Author.

<sup>8</sup> See National Council on Crime and Delinquency (2010) at endnote 7, p. 8. The initial risk scale was based on an outcome measure of weighted factors that included rules violations, arrests, misdemeanor convictions, absconsions, felony convictions, and convictions for assaultive offenses. See Baird et al. (1979) at endnote 4, p. 40.

<sup>9</sup> See Baird et al. (1979) at endnote 4, p. 8.

<sup>10</sup> See Baird et al. (1979) at endnote 4, p. 8.

<sup>11</sup> See Baird et al. (1979) at endnote 4.

<sup>12</sup> Baird, C., & Neuenfeldt, D. (1990). *Improving correctional performance through better classification: The Client Management Classification System*. Madison, WI: National Council on Crime and Delinquency.

<sup>13</sup> See National Council on Crime and Delinquency (2010) at endnote 7.

<sup>14</sup> National Council on Crime and Delinquency. (n.d.). *CAIS Correctional Assessment and Intervention System (brochure)*. Madison, WI: Author.

<sup>15</sup> See p. 154 in Harris, P. (1994). Client management classification and prediction of probation outcome. *Crime and Delinquency*, 40, 154-174.

<sup>16</sup> See p. 22 in Hubbard, D. J., Travis, L. F., & Latessa, E. J. (2001). *Case classification in community corrections: A national survey of the state of the art*. Cincinnati: University of Cincinnati. The percentages are based on the 288 agencies using a risk and needs assessment instrument. The entire sample consisted of 385 agencies.

<sup>17</sup> Taxman, F. S., Cropsey, K. L., Young, D. W., & Wexler, H. (2007). Screening, assessment, and referral practices in adult correctional settings: A national perspective. *Criminal Justice and Behavior*, 34, 1216-1234.

<sup>18</sup> W. Ore, personal communication, December 24, 2012.

<sup>19</sup> See NCCD: CAIS website at <http://www.nccdglobal.org/assessment/corr>

[ectional-assessment-and-intervention-system-cais](#).

<sup>20</sup> See National Council on Crime and Delinquency (2010) at endnote 7, p. 2.

<sup>21</sup> C. Baird, personal communications, March 21 and July 24, 2012.

<sup>22</sup> See Baird et al. (1979) at endnote 4, Appendix A, pp. 39-44 for a description of the Wisconsin risk assessment instrument. Unless otherwise noted, this document is used as the source for the development description in the text.

<sup>23</sup> C. Baird, personal communication, July 24, 2012. Also see Baird et al. (1979) at endnote 4, p. 42.

<sup>24</sup> See Baird et al. (1979) at endnote 4, p. 10.

<sup>25</sup> See Baird et al. (1979) at endnote 4, p. 10.

<sup>26</sup> C. Baird, personal communications, March 21 and July 24, 2012.

<sup>27</sup> See Baird et al. (1979) at endnote 4, p. 11.

<sup>28</sup> See p. 5 in Jones, D., Johnson, S., Latessa, E., & Travis, L. (1999). Case classification in community corrections: Preliminary findings from a national survey. In *Topics in Community Corrections*, 4-8. Washington, DC: U.S. Department of Justice, National Institute of Corrections. Also see See Baird et al. (1979) at endnote 4, p. 12.

<sup>29</sup> See Baird et al. (1979) at endnote 4, p. 12.

<sup>30</sup> See Baird et al. (1979) at endnote 4, pp. 12-13.

<sup>31</sup> See Baird et al. (1979) at endnote 4, p. 12.

<sup>32</sup> See Baird et al. (1979) at endnote 4, p. 14.

<sup>33</sup> See Baird et al. (1979) at endnote 4, p. 47. Additional information provided by C. Baird, personal communication, July 24, 2012: "Some agencies use a matrix that allows them to emphasize the role of one instrument...usually the risk assessment."

<sup>34</sup> C. Baird, personal communication, July 24, 2012.

<sup>35</sup> C. Baird, personal communication, March 21, 2012.

<sup>36</sup> See Baird et al. (1979) at endnote 4, p. 18.

<sup>37</sup> See p. 257 in Lerner, K., Arling, G., & Baird, C. (1986). Client management classification strategies for case supervision. *Crime and*

## Appendix: RNA Instrument Profile for the CAIS

---

*Delinquency*, 32, 254–271. Information on the development of the CMC also is available on pp. 75–78 in National Institute of Corrections. (1981). *NIC technical assistance report: Model probation/parole management program*. Washington, DC: Author. Unless otherwise noted, the profile’s description of the CMC’s development is based on these two documents.

<sup>38</sup> The taxonomy is sometimes reported as having five supervision strategies because one strategy has a subcategory. The CAIS System Manual presents the taxonomy as five strategies. See National Council on Crime and Delinquency (2010) at endnote 7, p. 4 and p. 6. However, the CAIS demonstration report provided by the test developers presents a classification score for the original four strategies.

<sup>39</sup> The ability of an item to differentiate among the supervision strategies was tested using a chi-square analysis. See National Institute of Corrections (1981) at endnote 37, pp. 76–77.

<sup>40</sup> Items were given a rating of 1, 2, or 3. Reliability ratings for each were at least .75, .80, and .90, respectively. Chi square significance levels for an item’s ability to differentiate among supervision strategy groups were at least .05, .01, and .001, respectively. Thus an item weighted as 3 for a particular supervision strategy group had an interrater reliability of at least .9 and differentiated the supervision strategy group from the other groups at a significance level of .001 or higher. See National Institute of Corrections (1981) at endnote 37, p. 76.

<sup>41</sup> See Baird & Neuenfeldt (1990) at endnote 12. The construction and validation samples were the same: C. Baird, personal communication, July 29, 2014.

<sup>42</sup> See National Institute of Corrections (1981) at endnote 37, p. 76.

<sup>43</sup> See Lerner et al. (1986) at endnote 37, p. 258.

<sup>44</sup> Ore, W., & Baird, C. (2014, March). *Beyond risk and needs assessments*. Madison, WI: National Council on Crime and Delinquency.

<sup>45</sup> C. Baird, personal communications, March 21.

<sup>46</sup> See Ore & Baird (2014, March) at endnote 44, p. 7. Information also provided by C. Baird, personal communications, March 21.

<sup>47</sup> C. Baird, personal communications, March 21.

<sup>48</sup> See National Council on Crime and Delinquency (2010) at endnote 7, pp. 7–8.

<sup>49</sup> See National Council on Crime and Delinquency (2010) at endnote 7, pp. 7–8.

The description also is based on example CAIS demonstration reports from 2009. The reports were provided by Toni Aleman of the National Council on Crime and Delinquency, August 17, 2010.

<sup>50</sup> See National Council on Crime and Delinquency (2010) at endnote 7, pp. 45 and 65 for CAIS risk items. See Baird et al. (1979) at endnote 4, pp. 10–11 for WRN risk items.

<sup>51</sup> The table does not include the 11th item included in the original WRN instrument because it was included at the request of the Wisconsin Department of Corrections and not because of its predictive ability. The WRN items also are out of order to better compare the items across the two instruments.

<sup>52</sup> See National Council on Crime and Delinquency (2010) at endnote 7, p. 3.

<sup>53</sup> See National Council on Crime and Delinquency (2010) at endnote 7, p. 7.

<sup>54</sup> See National Council on Crime and Delinquency (2010) at endnote 7, p. 6.

<sup>55</sup> See NCCD: CAIS webpage at <http://www.nccdglobal.org/assessment/correctional-assessment-and-intervention-system-cais>.

<sup>56</sup> See National Council on Crime and Delinquency (2010) at endnote 7, p. 4.

<sup>57</sup> C. Baird, personal communication, March 21, 2012. The focus was on ensuring that certain categories “are being considered for every case by every worker, and that the

## Appendix: RNA Instrument Profile for the CAIS

---

ratings are done fairly consistently across the raters.... And so a lot of agencies either added some areas or may have deleted some areas, depending on what...input that they got from people within their agency. The needs instrument, other than looking at the inter-rater reliability, is not a research-based instrument.”

<sup>58</sup> Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology*, 34, 575-607. The effect size adjusted for sample size is  $r = .32$ . The 14 effect sizes were derived from various studies that included outcome measures of arrest, conviction, incarceration, parole violation and/or some combination. The authors referred to “risk scales” (p. 585) when describing the instruments they examined; there is no indication that effect sizes also were calculated for the Wisconsin needs scale.

<sup>59</sup> At least one of the studies included in the meta-analysis examined the predictive validity of the risk assessment instrument without the assaultive factor and reported a correlation of  $r=.17$  with recidivism. See Wright, K. N., Clear, T. R., & Dickson, P. (1984). Universal applicability of probation risk-assessment instruments. *Criminology*, 22, 113-134. As a comparison, Robinson and Porporino did include the assaultive factor in their study, giving it a weight of 15 points. They reported a correlation of  $r=.21$  between the risk score and recidivism. See Robinson, D., & Porporino, F. J. (1989, May). *Validation of an adult offender classification system for Newfoundland and Labrador*. (Research report no. R-04). Ottawa, ON: Correctional Service of Canada.

<sup>60</sup> The data are summarized in Bonta, J. (1996). Risk-needs assessment and treatment. In A. T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 18-32). Thousand Oaks, CA: Sage. The author notes that a few modifications were made to the

instrument in 1986 but does not indicate the specific changes that were made.

<sup>61</sup> Eisenberg, M., Bryl, J., & Fabelo, T. (2009, August). *Validation of the Wisconsin Department of Corrections risk assessment instrument*. New York: Council of State Governments Justice Center. The correlation decreased to .18 with the assaultive factor included on the instrument.

<sup>62</sup> Henderson, H. & Miller, H. (2013). The (twice) failure of the Wisconsin Risk Need Assessment in a sample of probationers. *Criminal Justice Policy Review*, 24, 199-221. Wisconsin test developer Baird (C. Baird, personal communication, July 24, 2012) criticized the study for using a “substantially flawed outcome measure” and a “highly selective and limited sample” as well as misrepresenting prior work on the Wisconsin Risk and Needs Assessment.

<sup>63</sup> Latessa, E., Smith, P., Lemke, R., Makarios, M., & Lowenkamp, C. (2009). *Creation and validation of the Ohio risk assessment system: Final report*. Cincinnati, OH: University of Cincinnati Center for Criminal Justice Research.

<sup>64</sup> C. Baird, personal communication, January 27, 2012. In another communication, Baird further explained that “during the course of supervision, very high percentages of cases move to lower risk levels over time. The reclassification scale shifts emphasis from prior history items to factors that reflect behavior since the last assessment” (March 13, 2012).

<sup>65</sup> For the Nevada report, see Wagner, D., & Oremus, K. (2009, June). *Nevada Department of Public Safety Division of Parole and Probation risk and needs assessment validation*. Madison, WI: National Council on Crime and Delinquency. For the Wisconsin report, see Eisenberg et al. (2009, August) at endnote 61. For the Orange County report, see Eisenberg, M., Fabelo, T., & Tyler, J. (2011, October). *Validation of the Orange County California Probation Department risk assessment instrument*:

## Appendix: RNA Instrument Profile for the CAIS

---

*Final report.* New York: Justice Center, The Council of State Governments. For the Travis County report, see Bryl, J., Fabelo, T., & Nagy, G. (2006, August). *Travis community impact supervision: Guiding justice decisions with risk assessment instruments.*

Washington, DC: The JFA Institute.

<sup>66</sup> See Eisenberg et al. (2011, October) at endnote 65, pp. 40-46.

<sup>67</sup> For example, Robinson and Porporino (1989, May, at endnote 59) found a correlation of  $r=.14$  between the needs score and recidivism for a sample of 200 probation cases in Canada. They identified three needs items (interpersonal relationships, companions, and drug involvement) as significantly differentiating recidivists and non-recidivists (see Appendix C in report). While recidivism was related to both the risk and needs scales, neither scale differentiated well between medium and high risk offenders. This likely was due, in part, to the low base rate of recidivism (10.5%) for the entire sample. Bonta (1996, at endnote 60) found a slightly modified version of the needs assessment to be predictive for Manitoba probationers. Across a seven year period, the correlations ranged from  $r=.10$  to  $r=.22$ . In another study of over 11,000 Nevada offenders, Wagner and Oremus (2009, June, at endnote 65, pp. 27-29) found that 8 of the 11 needs items had a significant relationship to recidivism. Henderson and Miller (2013 at endnote 62) found a correlation of  $r=.19$  and an AUC=.62 between the total needs score and rearrest among a sample of 194 probationers in Texas. Their study identified three items (employment, financial management, and drug problems) as significantly related to recidivism.

<sup>68</sup> See Lerner et al. (1986) at endnote 37. The study also was reported in a publication by the National Council on Crime and Delinquency. See Ore & Baird (2014, March) at endnote 44. It is not clear from the two summaries when the data actually were collected. The Lerner et. al article says the

study was undertaken in 1979 and present data provided by the Wisconsin Division of Corrections in 1983. The Ore and Baird report refer to the “Wisconsin Study, 1986” (p. 6) and indicate that outcomes were measured 18 months after admission to probation. The size of the original sample also is unknown. Lerner et al. report that “the Ns for each outcome category varied somewhat due to missing information at termination” (p. 268). The sample size of 422 was based on those for whom information was available on revocations/discharges, the outcome measure with the most complete information.

<sup>69</sup> Eisenberg, M., & Markley, G. (1987). Something works in community supervision. *Federal Probation*, 51, 28-32.

<sup>70</sup> McManus, R. F., Stagg, D. I., & McDuffie, C. R. (1988). CMC as an effective supervision tool: The South Carolina perspective. *Perspectives, Summer*, 30-34.

<sup>71</sup> See Ore & Baird (2014, March) at endnote 44, p. 5. The study is also discussed in Harris, P. M., Gingerich, R., & Whittaker, T. A. (2004). The “effectiveness” of differential supervision. *Crime & Delinquency*, 50, 235-271.

<sup>72</sup> See Eisenberg et al. (2009, August) at endnote 61, p. 29 and Eisenberg et al. (2011, October) at endnote 65, p. 48.

<sup>73</sup> See Baird et al. (1979) at endnote 4, pp. 15-17.

<sup>74</sup> See National Institute of Corrections (1981) at endnote 37, p. 77. The report indicates that 59 CMC items had an inter-rater reliability of .9 or better, 70 items .8 or better, 97 items .7 or better, and 5 items slightly less than .7 (p. 76). Reportedly (C. Baird, personal communication, July 29, 2014), the number of items is based on the number of “forced-choice” options in the CMC and not the number of questions. Thus the number of items for which reliabilities were reported exceeds the number of questions on the instrument. At least three

## Appendix: RNA Instrument Profile for the CAIS

---

of the five raters assessed each of the 250 offenders in the sample.

<sup>75</sup> See Bryl et al. (2006, August) at endnote 65; Eisenberg et al. (2009, August) at endnote 61; Eisenberg et al. (2011, October) at endnote 65; and Wagner & Oremus (2009, July) at endnote 65. Eisenberg et al. (2009, August) also reported a correlation of  $r = -.073$ , “indicating a weak correlation between gender and new offense” (p. 23).

<sup>76</sup> See National Council on Crime and Delinquency (n.d.) at endnote 14. Also see NCCD CAIS website at <http://www.nccdglobal.org/assessment/correctional-assessment-and-intervention-system-cais>.

<sup>77</sup> See Eisenberg et al. (2009, August) at endnote 61; Eisenberg et al. (2011, October) at endnote 65; and Wagner & Oremus (2009, July) at endnote 65. Eisenberg et al. (2009, August) also reported a correlation of  $r = .05$  between race/ethnicity and new offense” (p. 25).

<sup>78</sup> See AutoMon, Assessments Management website page at <http://www.automon.com/solutions/criminal-justice/assessments>.

<sup>79</sup> C. Baird, personal communication, March 21, 2012.

<sup>80</sup> See National Council on Crime and Delinquency (n.d.) at endnote 14.

<sup>81</sup> C. Baird, personal communication, July 24, 2012.

<sup>82</sup> C. Baird, personal communication, July 24, 2012.

<sup>83</sup> C. Baird, personal communication, July 24, 2012.

<sup>84</sup> See Ore & Baird (2014, March) at endnote 44.

<sup>85</sup> See National Council on Crime and Delinquency (2010) at endnote 7.

<sup>86</sup> See National Council on Crime and Delinquency (2010) at endnote 7, p. 47.

<sup>87</sup> C. Baird, personal communication, July 24, 2012.

<sup>88</sup> See National Council on Crime and Delinquency (n.d.) at endnote 14.

# Correctional Offender Management Profile for Alternative Sanctions (COMPAS)

---

## COMPAS GLOSSARY OF TERMS

---

<b>Risk</b>	COMPAS distinguishes between risk scales, which are designed to measure the likelihood that an offender will recidivate, and needs scales, which are designed to capture information about offender needs that can be used to inform case plans and identify target criminogenic thoughts and behaviors for treatment intervention. <sup>1</sup> The authors “believe risk scales designed to predict risk should be dynamic (composed of dynamic criminogenic needs) so that one can measure changes in risk of recidivism over time.” <sup>2</sup>
<b>Static risk</b>	Authors indicate that these are “historical factors” (e.g., age at first arrest). <sup>3</sup>
<b>Dynamic risk</b>	Authors indicate that these are “criminogenic factors” (e.g., employment status, level of substance abuse). <sup>4</sup>
<b>Needs</b>	Offender needs are individual factors about the offender that, in the aggregate, have a demonstrated relationship with recidivism but that can be changed. <sup>5</sup> Included are factors such as criminal thinking, education, employment, substance abuse, residential stability and other aspects of the “person-in-environment” which guide individualized decisions in case planning. <sup>6</sup>
<b>Responsivity</b>	Responsivity refers to the principle that people respond differently to different treatment approaches. This recognizes that “the wrong treatment may make things worse and creates a need for careful matching of people to specific treatments.” <sup>7</sup> Officers who create the offender’s case plan must pay attention to responsivity issues at the intake assessment, as they capture information about the offender’s ability and readiness to make the changes to reduce their future likelihood of recidivating.
<b>Protective factors</b>	Protective factors are discussed as strengths (see below). <sup>8</sup>
<b>Strengths</b>	Offender factors (e.g., supportive families, educational and vocational strengths, stable residences in safe areas, social supports) that have shown empirical support for potential risk reduction and protecting individuals’ from the full impact of criminogenic needs. <sup>9</sup>
<b>Recidivism</b>	General recidivism refers to any new arrest within two years of the COMPAS assessment. <sup>10</sup>

---

## Appendix: RNA Instrument Profile for the COMPAS

---

### HISTORY & CURRENT USE.

---

**Creation.** COMPAS was initially developed in 1998 by the Northpointe Institute for Public Management. The instrument has since undergone several iterations of revisions and was last updated based on a national sample of 30,000 imprisoned and community-based offenders for whom COMPAS assessments were conducted between January 2004 and November 2005.<sup>11</sup> The current version of COMPAS has norms available for eight groups: male or female prison, jail, probation, or composite groups of offenders.<sup>12</sup>

**Current use.** COMPAS has been utilized by the California Department of Corrections and Rehabilitation, including probation departments in San Diego, San Francisco, Tulare, San Bernardino, and Riverside counties; Michigan Department of Corrections; New Mexico Corrections Department; New York State Department of Corrections and Community Supervision; South Carolina Department of Corrections; Wisconsin Department of Corrections; and Wyoming Department of Corrections.<sup>13</sup>

### DEVELOPMENT.

---

**Instrument purpose.** COMPAS is an automated, fourth generation risk and needs assessment instrument and case planning system.<sup>14</sup> It was “designed to help criminal justice practitioners determine the placement, supervision, and case-management of offenders in community and secure settings.”<sup>15</sup>

The COMPAS tool was designed to be adaptable for different agency decisions from

pretrial to prison release. The entire COMPAS system contains 42 separate scales that may be selected and combined for use with various offender populations (jail, prison, parole, probation) and at different decision points in the criminal justice process (pretrial release, case management).<sup>16</sup> The vendor provides client agencies with the version of the COMPAS that matches their needs.<sup>17</sup> For this reason, the actual uses and content of the COMPAS can vary substantially between agencies and between research studies. This profile focuses on the General Recidivism Risk scale and other components relevant for use with a general community-based population of adult offenders.

### **Approach to instrument development.**

Developers of COMPAS were strongly influenced by the process used to develop an outcomes-based recidivism scale for England and Wales.<sup>18</sup> In selecting and developing risk and needs scales for the COMPAS system, Northpointe undertook a theory-guided design based upon established causal theories of crime such as low self-control theory, social learning theory, strain theory, social control theory, routine activities-opportunity theory, and a strengths and good lives perspective.<sup>19</sup> The COMPAS scales also include key offender risk and needs factors that have emerged from meta-analytic research, including the “central 8.”<sup>20</sup> All COMPAS scales are composed of items selected by instrument developers on the basis of not only their relevance to factors theoretically associated with criminal behavior but also their demonstrated statistical relationship with those constructs.<sup>21</sup>

## Appendix: RNA Instrument Profile for the COMPAS

COMPAS distinguishes between risk scales and needs scales.<sup>22</sup> In the development of the risk scales, researchers prioritized the use of a limited set of items (parsimony) and the ability of risk scores to predict recidivism (predictive validity).<sup>23</sup> The General Recidivism Risk scale, was statistically derived based on data from a sample of presentence investigation and probation intake cases in 2002 to predict any offense (misdemeanor or felony) arrest within two years of the offender's COMPAS administration date.<sup>24</sup>

The needs scales (e.g., criminal thinking, education, employment, substance abuse, residential stability) capture and describe factors about the individual offender that have been found in the extant literature to be associated with criminal behavior.<sup>25</sup> These need areas are not all used in the calculation of offender recidivism risk; rather, they represent potential targets for treatment intervention to be used by the supervising officer to inform case planning efforts.<sup>26</sup>

COMPAS scales also include a mixture of both dynamic (e.g., level of substance abuse) and static items (e.g., age at first offense) to permit measurement of change over time. Although the exact items and proportion of static versus dynamic items may vary by scale and depending on the version of COMPAS used, over 50 percent of the items in COMPAS are dynamic.<sup>27</sup>

### CONTENT.

**Structure.** As indicated earlier, the exact structure of the COMPAS will vary by client agency. The entire COMPAS system contains 42 scales, including 4 offender recidivism

risk scales (e.g., General Recidivism Risk), 1 short 5-item recidivism risk screen scale, 19 gender-neutral "criminogenic need scales" to identify factors about the individual offender that are associated with criminal behavior in the larger population, 16 women-specific needs scales, and 2 validity scales.<sup>28</sup> The number of questions for each scale varies.<sup>29</sup>

**Items and domains.** The COMPAS Core Assessment includes 135 items that are combined into various risk and need scales.<sup>30</sup> The primary risk items within the General Recidivism Risk scale address prior criminal history, criminal associates, drug involvement, and early indicators of juvenile delinquency problems.<sup>31</sup> The 19 criminogenic need scales are organized into five overarching areas as described in Table 1.<sup>32</sup>

**Table 1. COMPAS Needs Scales**

Area	Scale
Criminal Involvement	<ul style="list-style-type: none"><li>• Criminal Involvement</li><li>• History of Non-Compliance</li><li>• History of Violence</li><li>• Current Violence</li></ul>
Relationships/Lifestyle	<ul style="list-style-type: none"><li>• Criminal Associates/Peers</li><li>• Criminal Opportunity</li><li>• Leisure/Recreation</li><li>• Social Isolation</li><li>• Substance Abuse</li></ul>
Personality/Attitudes	<ul style="list-style-type: none"><li>• Criminal Personality</li><li>• Criminal Thinking Self Report</li><li>• Cognitive Behavioral</li></ul>
Family	<ul style="list-style-type: none"><li>• Family Criminality</li><li>• Socialization Failure</li></ul>
Social Exclusion	<ul style="list-style-type: none"><li>• Financial</li><li>• Vocational/ Education</li><li>• Social Environment</li><li>• Residential Instability</li><li>• Social Adjustment</li></ul>

## Appendix: RNA Instrument Profile for the COMPAS

---

The specific items for each COMPAS scale are available from Northpointe.<sup>33</sup>

**Reporting and cutoffs.** The COMPAS software suite produces an individual assessment report to display each offender's results from the assessment tool in chart form.<sup>34</sup> Within the software application, raw scores are transformed into deciles, and each decile score is then used to determine the level of risk probability (deciles of 1-4 = low risk, 5-7 = medium risk, and 8-10 = high risk).<sup>35</sup> Cutoff scores for need scales vary by the scale with most falling into 1-5 = unlikely, 6-7 = probable, and 8-10 = highly probable.<sup>36</sup>

The decile scores and cutoffs are based upon a comparison of offender characteristics to a representative criminal population (i.e., a norming group). The norming group includes subpopulations of people from prison, jail, or probation.<sup>37</sup> Each agency has the ability to select a norming group that is most appropriate for its population of interest. For example, a probation agency might select the available probation sample as their norming group. Also, COMPAS can make use of separate norms for males and females to allow for gender-specific calibrations.

The assessment report chart of risk and needs scale results is accompanied by a narrative summary of the offender's assessment results. This document includes for each criminogenic need area a written description of the offender's need scale results, a statement from the interviewer, and a written description of associated treatment implications. Current charge and criminal history information are also presented.

COMPAS scales are also linked to specific "sets" of relevant treatment interventions and goals. These linkages are embedded within the COMPAS software and are offered as dropdown lists in the case plan section of the automated report. The lists of programs are based primarily on national evaluation research findings and the broader research literature with an emphasis on cognitive behavioral interventions, while simultaneously excluding programs shown to be ineffective by current evaluation research.<sup>38</sup> Program lists can be modified by client users based upon local knowledge of program effectiveness.

### INSTRUMENT RELIABILITY AND VALIDITY.

---

**Populations studied.** A number of internal and external validation studies have been conducted on COMPAS. These studies have focused on the use of the tool by the Michigan Department of Corrections,<sup>39</sup> New York State Division of Parole,<sup>40</sup> New York State Division of Probation and Correctional Alternatives,<sup>41</sup> and California Department of Corrections and Rehabilitation.<sup>42</sup> When implementing COMPAS in a new jurisdiction, Northpointe researchers typically incorporate an outcomes study with at least a year of follow-up for an initial analysis.<sup>43</sup>

Brennan, Dieterich and Ehret report that the COMPAS General Recidivism Risk scale also has been validated internally by Northpointe using "multi-year prospective outcome studies in new samples as well as for different racial/ethnic and gender groups across different state systems."<sup>44</sup> However, no comprehensive research publication of

## Appendix: RNA Instrument Profile for the COMPAS

---

these studies is publicly available at this time.

**Predictive validity.** The predictive validity of the COMPAS General Recidivism Risk scale has been examined in multiple internal pilot tests and outcome studies. Test developers report predictive validity Area Under the Curve (AUC) values ranging from .66 to .73 for any arrest.<sup>45</sup> An independent validation conducted on a sample of California parolees by Farabee and his colleagues reported an AUC of .70 for predicting any arrest within two years of being released from prison.<sup>46</sup> A study of 57 New York state probation departments using the COMPAS reported an AUC of .71 for predicting rearrest within two years among a sample of offenders admitted to probation in 2009.<sup>47</sup>

The Northpointe Practitioners Guide to COMPAS also reports on two studies examining the predictive validity of the COMPAS needs scales.<sup>48</sup> The first reports correlations ranging from  $r = -.07$  to  $r = .28$  and AUC values ranging from .51 to .63 across the 19 scales. The second study reports correlations ranging from  $r = -.16$  to  $r = .27$  and AUC values ranging from .50 to .66 across 18 of the scales.

**Reliability.** The test developers report average alpha scores measuring internal consistency of  $r = .70$  in a study of California prisoners and  $r = .73$  in a study of San Bernardino probationers.<sup>49</sup> They also report the alpha values for a combined sample of 47,679 males from California and Michigan Departments of Corrections ranging from  $r = .53$  to  $r = .86$ .<sup>50</sup>

With regard to test-retest reliability, Farabee and his colleagues reported correlations for COMPAS scales that ranged from .7 to 1.00 with an overall average score of .88, indicating that different assessment administrators provide consistent scoring of scale items.<sup>51</sup>

**Potential for bias.** The test developers report that they excluded all items that had any mention of racial, gender, religious or national origin issues in the assessment. They also report that the COMPAS scales show no systematic differences by race and gender on tests of internal consistency (Cronbach's alpha).<sup>52</sup>

- **GENDER.** COMPAS has gender-specific norm groups—female offender scores are compared to the scores of other females. Test developer Brennan and his colleagues report the predictive validity of COMPAS results did not differ significantly between men and women.<sup>53</sup> The test developers also report that COMPAS also now includes the new gender-responsive assessment designed and validated by Van Voorhis and colleagues at the University of Cincinnati.<sup>54</sup>
- **RACE.** COMPAS developers report finding very little variation in predictive validity between racial/ethnic groups.<sup>55</sup> One independent study came to a different conclusion about the predictive validity of the tool with minority offenders, concluding that the tool is only valid for use with Caucasians.<sup>56</sup> Northpointe researchers, however, argue that the sample size and base rates of

## Appendix: RNA Instrument Profile for the COMPAS

---

offending in the study were insufficient to address the question.<sup>57</sup>

**Independent validation.** A few independent evaluations of COMPAS have been conducted with mixed findings. Farabee and colleagues examined 91,334 parolees in California who had been assessed with COMPAS prior to release and had been on parole for at least one year. They concluded that the COMPAS had high test-retest reliability and acceptable predictive validity for the general recidivism risk scale.<sup>58</sup> As noted above, Fass and colleagues examined the predictive validity of the COMPAS using a male cohort of offenders released into the community from New Jersey prisons between 1999 and 2002, with a post-release outcome period of twelve months and found the COMPAS most predictive of Caucasian recidivism and least predictive of African American recidivism.<sup>59</sup> A third review by Skeem and Loudon examined the COMPAS based upon a synthesis of three extant reports.<sup>60</sup> They concluded that the COMPAS is relatively easy for professionals to apply and has internal consistency reliability. The authors concluded that there was no sound evidence to indicate predictive validity, construct/content validity, or high inter-rater reliability of the COMPAS.<sup>61</sup> Northpointe researchers contend that Skeem and Loudon's conclusions are invalid because their review was based on ongoing outcome studies with preliminary and incomplete data.<sup>62</sup>

### PRACTICAL CONSIDERATIONS.

---

**Vendor and instrument cost.** The COMPAS is a proprietary tool offered by

Northpointe, Inc. For more information on the instrument and software packages available, refer to their website at [www.northpointeinc.com](http://www.northpointeinc.com).<sup>63</sup>

**Menu of other services.** COMPAS offers a wide array of services, training, and technical assistance to support implementation.<sup>64</sup>

- **IT SERVICES.** The COMPAS software is scalable depending on client decision support needs and includes a user-configurable case planning module that is prepopulated with offender needs assessment results. IT customization services are available from Northpointe, and clients also can opt to have the application hosted on Northpointe's system rather than integrated into the client's system.<sup>65</sup>
- **VALIDATION SERVICES.** Northpointe offers clients an array of research services, including local validation research studies. The costs will vary depending on sample size, length of outcome, and the scope of the study (e.g., overall predictive validity or breakdowns by gender, race, ethnicity, and/or other factors).<sup>66</sup>
- **REASSESSMENTS.** Offender reassessment is built into the software to allow direct comparisons of offender profiles across time.<sup>67</sup> Northpointe leaves the decision to reassess to the discretion of the agency based on factors such as case management goals and objectives, length of time the offender is under supervision, staff resources, and so forth. If an agency opts to reassess, Northpointe suggests that it be conducted at least 8 to 12 months after the initial COMPAS

## Appendix: RNA Instrument Profile for the COMPAS

---

assessment to better measure true offender changes.<sup>68</sup>

- **USER TRAINING.** The COMPAS standard two-day training is mandatory and is typically included in the purchase of the system.<sup>69</sup> Users are instructed on how to use the tool, interpret assessment results, and create case plans for offenders to address high-need areas.<sup>70</sup> Northpointe also offers additional training options, depending on an agency's needs.<sup>71</sup>
- **CUSTOMER SUPPORT.** Northpointe provides customer and technical support Monday through Friday, 8AM to 5PM ET.<sup>72</sup>

**User qualifications.** Any user of the tool must complete a two-day COMPAS user training.<sup>73</sup> The instrument can be used by those with limited computer experience and education.<sup>74</sup>

**Administration time.** Depending on the version of COMPAS selected for use by the client agency, the assessment may take between ten minutes to one-hour.<sup>75</sup>

In an independent survey study of Parolee Services administrators in California, test administrators reported taking an average of 39 minutes to administer the COMPAS re-entry assessment interview, 58 minutes reviewing an offender's file, and 24 minutes to enter the results into the database.<sup>76</sup>

**Modes of administration.** COMPAS relies upon three procedures to collect information. First, data are gathered from official records by a criminal justice professional. Second, a trained test administrator conducts a structured

interview with the offender. Third, offenders complete a self-reported paper and pencil questionnaire. Each data modality accounts for about one-third of the data collected.<sup>77</sup>

The entire COMPAS system is automated, but requires manual input of the raw data collected by the test administrator. In some instances the official criminal records can automatically populate the criminal history section of COMPAS, where the appropriate transfer software is present.

Northpointe also offers an Ad-Hoc Report Generator that allows for client customization of various management and monitoring reports. These reports can be exported into PDF format or excel, word, XML, or RFT for import into statistical packages for further analysis.<sup>78</sup>

**Quality assurance.** When adopting any offender assessment tool, jurisdictions must be prepared to ensure appropriate implementation and proper maintenance over time. Quality assurance recommendations and guidelines for the COMPAS follow.

- **OVERRIDE POLICY.** COMPAS designers expect staff to disagree with COMPAS in about ten percent of cases due to mitigating or aggravating circumstances.<sup>79</sup> Northpointe defines mitigating factors as those that "may excuse the offender, reduce the seriousness of the crime or raise the likelihood of a pro-social adjustment."<sup>80</sup> They define aggravating factors as "extraneous information that makes the offense more serious, more violent, or may appear to make the offender more culpable, more resistant to treatment,

## Appendix: RNA Instrument Profile for the COMPAS

---

and so forth.”<sup>81</sup> In these cases, staff is encouraged to use their professional judgment to override the scale results. It is suggested that staff document the override reason and make such reasons available to supervisory staff for monitoring.<sup>82</sup>

- **FIDELITY.** COMPAS’ multi-modal data collection is designed to promote assessment reliability and ensure corroboration of offender responses. In addition, COMPAS Core contains two scales designed to examine the validity of offender responses to self-report items. One of these scales tests the offender for extreme responses (the Lie Scale) and the other tests offender responses for consistency (the Random Responding Scale).<sup>83</sup> These scales were introduced as a means to detect when offenders deliberately provide false responses to self-report items, signaling to the test administrator that further scrutiny may be required.
- **INSTRUMENT REVALIDATION.** Northpointe encourages periodic local validation of the COMPAS, as frequently as every other year.<sup>84</sup> However, they note that they have not yet found any “statistically significant deviations” in local validations of the COMPAS from national norm group studies.<sup>85</sup>

### ENDNOTES

---

<sup>1</sup> Northpointe Institute for Public Management (2012). *Practitioner’s guide to COMPAS*. Traverse City, MI: Author.

<sup>2</sup> Northpointe (2012) at endnote 1, p. 15.

<sup>3</sup> See Northpointe (2012) at endnote 1, p. 1.

<sup>4</sup> See Northpointe (2012) at endnote 1, p. 1.

<sup>5</sup> See Northpointe (2012) at endnote 1, p. 17.

<sup>6</sup> See Northpointe (2012) at endnote 1, p. 17.

<sup>7</sup> See Northpointe (2012) at endnote 1, p. 44.

<sup>8</sup> See p. 23 in Brennan, T. Dieterich, W., & Ehret, B. (2009). Evaluating the predictive validity of the COMPAS risk and needs assessment system. *Criminal Justice and Behavior*, 36, 21-40.

<sup>9</sup> See Brennan et al. (2009) at endnote 8, p. 23. Information also provided by T. Brennan, personal communication, August 15, 2012.

<sup>10</sup> See p. 15 in Northpointe Institute for Public Management. (2010). *Practitioner’s guide to COMPAS*. Traverse City, MI: Author.

<sup>11</sup> See Northpointe (2012) at endnote 1, p. 2.

<sup>12</sup> See Northpointe (2012) at endnote 1, p. 9.

<sup>13</sup> This list may not be exhaustive of all locations that have used or are using COMPAS. The list is drawn from Brennan, T., Dieterich, W., & Ehret, B. (2007). *Research synthesis: Reliability and validity of COMPAS*. Traverse City, MI: Northpointe; Northpointe (2012) at endnote 1, p. 15; the Wisconsin Department of Corrections website at <http://doc.wi.gov/about/doc-overview/office-of-the-secretary/office-of-reentry/compas-assessment-tool>; and W. Dieterich, personal communication, June 16, 2014.

<sup>14</sup> See Northpointe (2012) at endnote 1, p. 1.

<sup>15</sup> See p. 3 in Farabee, D., S. Zhang, R. Roberts, and J. Yang (2010). *COMPAS validation study: Final report* (Tech. Rep.). Los Angeles: Semel Institute for Neuroscience and Human Behavior University of California.

<sup>16</sup> See Northpointe (2012) at endnote 1, p. 2.

<sup>17</sup> See Northpointe (2012) at endnote 1, p. 2. For example, the specialized REENTRY COMPAS was designed for use with longer-term prison inmates, who may have distinct needs following a lengthy period of incarceration (e.g., social support, housing needs). T. Brennan, personal communication, August 15, 2012.

<sup>18</sup> See Northpointe (2012) at endnote 1, p. 15.

## Appendix: RNA Instrument Profile for the COMPAS

---

<sup>19</sup> See Northpointe (2012) at endnote 1, pp. 6-8 and Brennan, et al. (2009) at endnote 8, p. 3.

<sup>20</sup> See Brennan, et al. (2009) at endnote 8, p. 3. The central (or big) eight factors include: antisocial attitudes, antisocial associates, a history of antisocial behavior, antisocial personality pattern, problematic circumstances at home, difficulties at work or school, problems with leisure activities, and substance abuse. See Andrews, D., J. Bonta, and S. Wormith (2006). The recent past and near future of risk and/or need assessment. *Crime and Delinquency*, 7-27.

<sup>21</sup> See Northpointe (2012) at endnote 1, p. 15, 17. Also see p. 16 in Brennan, T., Dieterich, W., Breitenbach, M., & Mattson, B. (2009). *A response to "Assessment of evidence on the quality of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)."*  Traverse City, MI: Author.

<sup>22</sup> See Northpointe (2012) at endnote 1, p. 14.

<sup>23</sup> See Northpointe (2010) at endnote 10, p. 15.

<sup>24</sup> See Northpointe (2010) at endnote 10, pp. 15, 24. The risk scale was derived from a regression model with bootstrap validation. Though not the subject of this profile, the other COMPAS risk scales (Violent Recidivism Risk, Pretrial Release Risk or Failure-to-Appear Risk, and Community Non-Compliance Risk) were similarly developed using different construction samples of offenders. The Violent Recidivism Risk Scale was developed in 2006 based on a sample of presentence investigation and probation intake cases to predict violent offenses within two years of intake. The Pretrial Release Risk Scale was developed in 2010 based upon felony defendants assessed with COMPAS in Kent County, Michigan. The scale is used to predict failure to appear and new felony arrests. See Northpointe (2010) at endnote 10, pp. 15-16. The Community Non-Compliance Risk Scale was designed to predict community non-compliance (likelihood of offender technical

violations) for probation and parole agencies. It is currently undergoing revision by the instrument developers. T. Brennan, personal communication, August 15, 2012.

<sup>25</sup> See Northpointe (2012) at endnote 1, p. 17.

<sup>26</sup> Research developers applied psychometric procedures such as factor analysis to maximize scale coherence, or logical consistency; see Brennan, T., Dieterich, W., & Ehret, B. (2007). *Research synthesis: Reliability and validity of COMPAS*. Traverse City, MI: Northpointe Institute for Public Management.

<sup>27</sup> T. Brennan, personal communication, August 15, 2012.

<sup>28</sup> See Northpointe (2012) at endnote 1, p. 2. Though not discussed in the *Practitioners Guide to COMPAS*, Community Non-compliance scale, currently under revision, is one of the four risk scales; T. Brennan, personal communication, August 15, 2012. The 5-item recidivism risk screen scale is a new addition to the system; W. Dieterich, personal communication, June 16, 2014.

<sup>29</sup> Northpointe Institute for Public Management. (2008). *Measurement & treatment implications of adult COMPAS scales*. Traverse City, MI: Author.

<sup>30</sup> L. Morris, personal communication, August 11, 2010.

<sup>31</sup> See Northpointe (2012) at endnote 1, p. 24.

<sup>32</sup> L. Morris, personal communication, August 11, 2010.

<sup>33</sup> See Northpointe (2008) at endnote 29.

<sup>34</sup> See example in COMPAS flyer: Northpointe Institute for Public Management. (n.d.). *COMPAS/core*. Williamsburg, MI: Author (available [http://www.northpointeinc.com/files/downloads/COMPAS\\_Core\\_Flyer\\_Front.pdf](http://www.northpointeinc.com/files/downloads/COMPAS_Core_Flyer_Front.pdf) and [http://www.northpointeinc.com/files/downloads/COMPAS\\_Core\\_Flyer\\_Back.pdf](http://www.northpointeinc.com/files/downloads/COMPAS_Core_Flyer_Back.pdf).)

<sup>35</sup> See Northpointe (2012) at endnote 1, pp. 9-11. Decile scores are provided as a percentile ranking system that compares the offender's raw score on a particular scale with the scores from a normative group. For example,

## Appendix: RNA Instrument Profile for the COMPAS

---

a decile score of “1” on a particular scale indicates that the offender’s score falls in the lowest 10% of all offenders in the normative group.

<sup>36</sup> See Northpointe (2012) at endnote 1, p. 11.

<sup>37</sup> See Northpointe (2012) at endnote 1, p. 5.

<sup>38</sup> T. Brennan, personal communication, August 15, 2012.

<sup>39</sup> Brennan, T. & Dieterich, W. (2008). *Michigan Department of Corrections Core COMPAS pilot study: One-year follow-up* (Tech. Rep.). Traverse City, MI: Northpointe. Dieterich, W., Brennan, T. & Oliver, W. (2011). *Predictive validity of the Reentry COMPAS Core risk scales: A probation outcomes study conducted for the Michigan Department of Correction* (Tech Rep.). Traverse City, MI: Northpointe.

<sup>40</sup> Brennan, T., Dieterich, W., & Breitenbach, M. (2008). *New York State Division of Parole COMPAS Reentry pilot study: Two-year follow-up: Updated predictive models*. (Tech Rep.). Traverse City, MI: Northpointe.

<sup>41</sup> Brennan, T., & Dieterich, W. (2009). *Testing the predictive validity of the DPCA COMPAS risk scales: Phase I* (Tech. Rep.). Traverse City, MI: Northpointe.

<sup>42</sup> See Farabee et al. (2010) at endnote 15. Also see review by Skeem, J., & Loudon, J. E. (2007). *Assessment of evidence on the quality of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)*. Davis, CA: Center for Public Policy Research.

<sup>43</sup> See Brennan et al. (2009) at endnote 21, p. 8.

<sup>44</sup> See Brennan et al. (2009) at endnote 21, p. 8.

<sup>45</sup> See Brennan et al. (2009) at endnote 8, p. 30 and Northpointe (2012) at endnote 1, p. 16. The AUC’s reported for the New York probation study (n=2,328) differed in the two references (.66 for the former and .71 for the latter) because of different analyses: W. Dieterich, personal communication, June 7, 2014.

<sup>46</sup> See Farabee (2010) at endnote 15, p. 24.

The AUC value for predicting a rearrest for a violent offense within two years was .65.

<sup>47</sup> Lansing, S. (2012, September). *New York State COMPAS-probation risk and need assessment study: Examining the recidivism scale’s effectiveness and predictive accuracy*. Albany, NY: Division of Criminal Justice Services, Office of Justice Research and Performance.

<sup>48</sup> Northpointe Institute for Public Management (2013). *Practitioner’s guide to COMPAS*. Traverse City, MI: Author. The authors report that the data used to compute the predictive validity of the needs scales come from the study by Farabee et al. (2010) at endnote 15 and Brennan et al. (2009) at endnote 8. The predictive validity scales were not reported in the original articles. The Farabee et al. study included 23,635 offenders who were followed for two years after their release from prison. The outcome measure was any arrest. The Brennan et al. study included 2,328 probationers who were followed for one year after intake. The outcome measure was felony arrest.

<sup>49</sup> See Northpointe (2012) at endnote 1, p. 21.

<sup>50</sup> See Northpointe (2012) at endnote 1, p. 22.

<sup>51</sup> Reliability scores are based on all scales without a distinction being made between the General Recidivism and the Violent Recidivism scales. See Farabee et al. (2010) at endnote 15, pp. 13-14.

<sup>52</sup> T. Brennan, personal communication, August 15, 2012.

<sup>53</sup> See Brennan et al. (2009) at endnote 8, p. 33.

<sup>54</sup> See, for example, Van Voorhis, P., Wright, E. M., Salisbury, E., & Bauman, A. (2010). Women’s risk factors and their contributions to existing risk/needs assessment: The current status of a gender-responsive supplement. *Criminal Justice and Behavior*, 37, 261-288. Inclusion in COMPAS noted in T. Brennan, personal communication, August 15, 2012.

<sup>55</sup> See Brennan et al. (2009) at endnote 8, p. 33.

## Appendix: RNA Instrument Profile for the COMPAS

---

<sup>56</sup> Fass, A. W., Heilbrun, K., DeMatteo, D., & Fretz, R. (2008). The LSI-R and the COMPAS: Validation data on two risk-needs tools. *Criminal Justice and Behavior*, 35, 1095-1108. Fass and colleagues examined the predictive validity of the COMPAS for different racial and ethnic groups and reported an AUC value of .81 for male Caucasians, .48 for male African Americans, and .67 for male Hispanics.

<sup>57</sup> See Brennan et al. (2009) at endnote 8, p. 33. Also see Northpointe (2012) at endnote 1, p. 16.

<sup>58</sup> See Farabee et al. (2010) at endnote 15, pp. 3-4.

<sup>59</sup> See Fass et al. (2008) at endnote 56, p. 7.

<sup>60</sup> See Skeem, J., & Louden, J. E. (2007) at endnote 42.

<sup>61</sup> See Skeem, J., & Louden, J. E. (2007) at endnote 42, p. 6.

<sup>62</sup> See Brennan et al. (2009) at endnote 21.

<sup>63</sup> Additional contact is available with Northpointe representatives by phone at 888-221-4615 or by email at [info@npipm.com](mailto:info@npipm.com).

<sup>64</sup> See Northpointe (n.d.) at endnote 34.

<sup>65</sup> See Northpointe's Technology web page at <http://www.northpointeinc.com/services/technology>. Also see pp. 9-11 in Northpointe Institute for Public Management (2012). *COMPAS risk & need assessment system: Selected questions posed by inquiring agencies*. Traverse City, MI: Author (available [http://www.northpointeinc.com/files/downloads/FAQ\\_Document.pdf](http://www.northpointeinc.com/files/downloads/FAQ_Document.pdf)).

<sup>66</sup> See Northpointe Institute for Public Management (2012) at endnote 65, pp. 6, 9. Information also provided by T. Brennan, personal communication, August 15, 2012.

<sup>67</sup> See Northpointe Institute for Public Management (2012) at endnote 65, p. 8. Information also provided by T. Brennan, personal communication, August 15, 2012.

<sup>68</sup> See Northpointe Institute for Public Management (2012) at endnote 65, p. 8.

<sup>69</sup> T. Brennan, personal communication, August 15, 2012.

<sup>70</sup> See Northpointe Institute for Public Management (2012) at endnote 65, p. 7.

<sup>71</sup> See Northpointe's Training web page at <http://www.northpointeinc.com/services/training> and Northpointe (2012) at endnote 65, pp. 7, 9.

<sup>72</sup> See Northpointe Institute for Public Management (2012) at endnote 65, p. 9.

<sup>73</sup> T. Brennan, personal communication, August 15, 2012.

<sup>74</sup> See Northpointe Institute for Public Management (2012) at endnote 65, p. 7.

<sup>75</sup> See Northpointe Institute for Public Management (2012) at endnote 65, p. 1.

<sup>76</sup> See Farabee et al. (2010) at endnote 15.

<sup>77</sup> See Brennan et al. (2009) at endnote 21, pp. 21-22. Information also from T. Brennan, personal communication, August 15, 2012.

<sup>78</sup> See Northpointe's Ad Hoc Report Generator web page at <http://www.northpointeinc.com/products/ad-hoc-report-generator> and Northpointe (2012) at endnote 65, pp. 8, 10.

<sup>79</sup> See Northpointe (2012) at endnote 1, p. 28.

<sup>80</sup> See Northpointe Institute for Public Management (2012) at endnote 65, p. 6.

<sup>81</sup> See Northpointe Institute for Public Management (2012) at endnote 65, p. 6.

<sup>82</sup> See Northpointe Institute for Public Management (2012) at endnote 1, p. 28.

<sup>83</sup> See Northpointe (2012) at endnote 1, pp. 42-43.

<sup>84</sup> T. Brennan, personal communication, August 15, 2012.

<sup>85</sup> See Northpointe Institute for Public Management (2012) at endnote 65, p. 6.

# Level of Service Assessments: Level of Service Inventory-Revised (LSI-R) Level of Service/Case Management Inventory (LS/CMI)

---

## LS GLOSSARY OF TERMS

---

<b>Risk</b>	Risk factors refer “to characteristics of people and their circumstances that are associated with an increased chance of future criminal activity.” <sup>1</sup> An offender’s risk level is important to decisions of release, supervision, and the allocation of treatment resources. “According to the risk principle of case classification, more intensive services are best allocated to the higher-risk cases while low-risk cases have a low probability of recidivism even in the absence of treatment services.” <sup>2</sup> Level of risk is measured with both static and dynamic risk factors (see below).
<b>Static risk</b>	Static risk factors are fixed or stable offender characteristics and aspects of personal history, such as age of first offense, that are related to the risk of reoffending. <sup>3</sup>
<b>Dynamic risk</b>	Dynamic risk factors, also referred to as criminogenic need factors, refer to risk factors that can change (e.g., antisocial attitudes) and thus “suggest appropriate intermediate targets” for reducing recidivism. <sup>4</sup>
<b>Needs</b>	Authors differentiate between criminogenic needs – problematic circumstances related to the risk of reoffending (see dynamic risk, above) – and noncriminogenic needs – problematic circumstances (e.g., homelessness) not related to the risk of reoffending. <sup>5</sup>
<b>Responsivity</b>	Responsivity refers to delivering treatment programs consistent with an offender’s ability and learning style. General responsivity – using social learning and cognitive-behavioral principles to change behavior – is distinguished from specific responsivity – offender characteristics (e.g., cognitive development) that may affect an offender’s success in a program. <sup>6</sup> Responsivity characteristics are not necessarily related to risk, “but they should be considered, particularly in the planning of intervention strategies.” <sup>7</sup>
<b>Protective factors</b>	The authors use the terms “strength” (see below) and “protective” factors synonymously. <sup>8</sup>
<b>Strengths</b>	“Strengths refer to characteristics of people and their circumstances that are associated with reduced chances of criminal activity.” <sup>9</sup> Strengths “may serve as protective factors and actively reduce the chances of antisocial conduct.” <sup>10</sup>
<b>Recidivism</b>	Recidivism has been defined variously (e.g., new arrest, new conviction, new incarceration) across studies examining the LS instruments. <sup>11</sup>

---

# Appendix: RNA Instrument Profile for the LSI-R and LS/CMI

---

## HISTORY & CURRENT USE

---

**Creation.** In the late 1970s, Canadian psychologist Don Andrews consulted with the Ontario Ministry of Correctional Services to develop a convenient, standardized, and reasonably comprehensive tool to record offender attributes and help probation officers make decisions about the level of supervision an offender would need.<sup>12</sup> Initial versions of the tool were tested and refined by Andrews and colleague James Bonta and subsequently published as *The Level of Service Inventory-Revised (LSI-R)* in 1995.<sup>13</sup> This instrument is still available and widely used; however, in 2004, Andrews and Bonta, joined by colleague J. Stephen Wormith, published an updated version of the LSI-R, the *Level of Service/Case Management Inventory (LS/CMI)*, that includes additional sections designed to help with generating an offender's case plan and monitoring progress on its implementation.<sup>14</sup>

**Current use.** As of 2010, the LS instruments' developers report widespread use of the assessments, including jurisdictions in 23 states and Puerto Rico in America, 9 Canadian jurisdictions, and several other countries around the world.<sup>15</sup>

## DEVELOPMENT

---

**Instrument purpose.** The LS instruments are designed to help probation, parole, and other correctional officers identify areas of risk and needs that can be addressed with programming to reduce offender risk while applying the least restrictive and onerous supervision necessary for safety.<sup>16</sup> The instrument developers sought to make the

information used for risk and treatment decisions transparent and consistent across correctional officers.<sup>17</sup> In addition, the LS/CMI focuses on additional information relevant to case management, treatment planning, and service delivery.<sup>18</sup>

**Approach to instrument development.** Personality and social learning perspectives of criminal conduct, research on recidivism, and the professional opinions of probation officers guided the development of the LS instruments.<sup>19</sup> From these sources of information, a large list of potential items was generated and subsequently screened for redundancy, theoretical consistency, and predictive ability.<sup>20</sup> Through this process, the instrument's developers identified a set of items they thought predictive of recidivism *and* useful for case management and treatment planning. The latter purpose they considered crucial for helping probation officers identify intermediate targets of change. Thus the developers took a more comprehensive approach to item selection rather than identifying the minimum number of items most predictive of recidivism alone.<sup>21</sup>

Ottawa probation officers began scoring offenders on the LSI-VI version of the instrument, consisting of 58 items, in the summer of 1980. The first 598 offenders receiving the LSI-VI assessment served as the initial validation sample. This early evaluation demonstrated that LSI-VI scores were related to probation officers' risk decisions and in-program recidivism outcomes.<sup>22</sup> Following additional testing and refinement, the developers eventually published the 54-item LSI-Revised in 1995.<sup>23</sup>

## Appendix: RNA Instrument Profile for the LSI-R and LS/CMI

---

Canadian samples of 956 male offenders from two detention centers and a jail, and 1,414 female offenders from a medium security institution for adult women, serve as the norm (or reference) groups for assessing an offender's risk level.<sup>24</sup> Normative data from the U. S., added in 2003, consists of 23,721 male and females who are in community corrections or incarcerated.<sup>25</sup> The community offender sample—those on probation or parole—includes 4,240 individuals from seven samples in two midwestern states, one southern state, and one northeastern state.<sup>26</sup> Normative data are provided for male inmates, male community offenders, female inmates and female community offenders.

Beginning in 1994, the Ontario Ministry of Community Safety and Correctional Services initiated a review of the LSI-R to address users' concerns (e.g., validity of the LSI-R with specific types of offenders, the omission of strengths and noncriminogenic needs) and produce training materials that better linked LSI-R use with evidence-based correctional practices.<sup>27</sup> The review involved broad consultation with representatives of community and institutional corrections, research and training units, and a variety of related government offices and professional associations. This feedback led to the development of the LSI-Ontario Revision which was the foundation for the LS/CMI. In particular, the LS/CMI's manual and scoring instructions were modified for application to a wider range of jurisdictions.<sup>28</sup>

Differences between the LSI-R and the LS/CMI are the latter's greater focus on the central 8 factors identified in the research literature as most predictive of recidivism

and the elimination of items with no or very low correlation with recidivism in calculating the risk/need level.<sup>29</sup> In addition, new sections were added to the LS/CMI to sample case strengths, responsivity considerations, specific risk/need factors and noncriminogenic needs.<sup>30</sup>

The LS/CMI was developed based on the results of studies conducted by the Ontario Ministry of Community Safety and Correctional Services. The LS/CMI offers Canadian, U.S., United Kingdom, and Singaporean normative data as well as data on young offenders.<sup>31</sup> The U. S. normative data is based on 48,384 offenders from nine geographically diverse jurisdictions.<sup>32</sup> Like the LSI-R, norms are available for four groups: male inmates, male community offenders, female inmates and female community offenders.

### CONTENT

---

**Structure.** Both the LSI-R and the LS/CMI calculate a single risk and needs score. Items are scored or recoded as either yes (1) or no (0) and then summed for the total score. Both instruments include a profile form that easily converts the raw score to a percentile, based on the appropriate normative group.

The LS/CMI includes ten additional sections that gather data on factors that may influence an offender's behavior. These include:

- Specific Risk/Need Factors (Section 2),
- Prison Experience—Institutional Factors (Section 3),
- Other Client Issues (Section 4),

## Appendix: RNA Instrument Profile for the LSI-R and LS/CMI

- Special Responsivity Considerations (Section 5),
- Risk/Need Summary and Override (Section 6),
- Risk/Need Profile (Section 7),
- Program/Placement Decision (Section 8),
- Case Management Plan (Section 9),
- Progress Record (Section 10), and
- Discharge Summary (Section 11).<sup>33</sup>

These additional sections are not scored; they provide qualitative information important to supervision and treatment decisions.

**Items and domains.** The LSI-R consists of 54 items across 10 subcomponents, and the LS/CMI consists of 43 items across 8 subcomponents (see Table 1).<sup>34</sup> Both instruments include static and dynamic risk items.

**Table 1. LSI-R and LS/CMI Subcomponents**

Subcomponent	# of Items	
	LSI-R	LS/CMI
Criminal History	10	8
Education/Employment	10	9
Financial	2	
Family/Marital	4	4
Accommodation	3	
Leisure/Recreation	2	2
Companions	5	4
Alcohol/Drug Problems	9	8
Emotional/Personal	5	
Attitudes/Orientation	4	4*
Antisocial Pattern		4

\*Renamed Procriminal Attitude/Orientation in LS/CMI

The LS/CMI omits the financial and accommodation subcomponents of the LSI-

R. In addition, the LS/CMI has a new subcomponent called antisocial pattern which is comprised of some of the emotional/personal items on the LSI-R. The LS/CMI also allows the test administrator to indicate whether a subcomponent is considered a strength for the offender and thus could be used in case planning to help address other problem areas.

**Reporting risk levels.** The LSI-R groups offenders on probation into three levels of risk (minimum, medium, maximum) based on their overall score.<sup>35</sup> The LS/CMI groups offenders on probation into 5 levels of risk (very low, low, medium, high, very high) based on their overall scores.<sup>36</sup> The cutoff scores are the same for males and females.

The test developers suggest a range of total risk and needs scores to include in each risk level; however, they strongly recommend that jurisdictions develop their own classifications based on research and local considerations such as staff resources, tolerance for failure, and available security options.

Another step in interpreting the results of the assessment is to consider the offender's score on each subcomponent. Those subcomponents with higher scores indicate areas to address in the offender's case plan.<sup>37</sup> The LS/CMI includes "Section 7: Risk/Need Profile" in which each subcomponent score is transferred to a table that identifies the risk/need level for that subcomponent. Thus the offender is rated very low, low, medium, high, or very high on each subcomponent, too.<sup>38</sup> In addition, the LS/CMI suggests that any subcomponent designated a "strength" also should be considered in developing an offender's case plan.

### INSTRUMENT RELIABILITY AND VALIDITY.

---

**Populations studied.** As noted under the “Development” section, the normative groups for the LS instruments include males and females, adults and juveniles, individuals in a range of correctional settings, and individuals from several different countries. In addition, other researchers have studied the instruments in a variety of jurisdictions; some examples of these studies follow.

**Predictive validity.** A 2013 meta-analytic review of 30 years of research on the LS scales conducted by Olver and his colleagues found that, across 124 samples and a total of 130,833 adult and juvenile offenders from around the world, the LS total scores significantly predicted general community recidivism ( $r_w = .30$  and  $.29$  for fixed- and random-effects models, respectively).<sup>39</sup> Vose and her colleagues’ 2008 review of 47 studies involving adults, juveniles, males and females in a variety of correctional placements in the United States, Canada, and Europe found a statistically significant relationship between total LS score and recidivism in 81% of the studies and a positive relationship between LS scores and recidivism in 98% of the studies.<sup>40</sup> The correlations across studies examining new charges, re-arrest, reconviction, and reincarceration ranged from  $r = .06$  to  $r = .51$ .<sup>41</sup> A 1996 meta-analysis by Gendreau and his colleagues yielded a mean effect size of  $r = .35$ .<sup>42</sup> A second meta-analysis by Gendreau and his colleagues in 2002 resulted in a mean effect size of  $r = .37$ .<sup>43</sup> Based on these meta-analyses and an additional study by Hemphill and Hare; Andrews, Bonta, and

Wormith summarized the effect size of the LSI-R in a 2006 article as  $.36$  for predicting general recidivism.<sup>44</sup> In addition, studies using receiver operating characteristic analysis have reported areas under the curve (AUC) of  $.689$  for a sample of federal probationers,  $.644$  for a sample of Iowa probationers, and  $.652$  for a sample of Iowa parolees.<sup>45</sup>

It should be noted that most of these studies have focused on the predictive validity of the LSI-R. Because Section 1 of the LS/CMI is highly correlated with the LSI-R, the test developers believe the predictive validity of the LS/CMI is equal or better than the LSI-R.<sup>46</sup> At least one study confirms their belief, finding a correlation of  $r = .39$  between LS/CMI total risk scores and recidivism.<sup>47</sup>

**Dynamic predictive validity.** A few studies have examined whether changes in LSI-R scores over time are related to changes in recidivism rates. Andrews, Bonta, and Wormith report five studies indicating that changes in risk level at follow-up assessments were related to expected changes in subsequent recidivism rates.<sup>48</sup> For example, those whose risk scores increased from the first assessment to the second assessment had higher rates of recidivism than those whose scores remained low. However, one of the studies also found that after the first reassessment, additional reassessments added limited improvement to overall predictive validity, suggesting that additional research is needed to fully understand when and how often reassessment is warranted.<sup>49</sup>

**Reliability.** Reliability values for the LS instruments are available for the consistency between raters’ scores, the stability of an

---

## Appendix: RNA Instrument Profile for the LSI-R and LS/CMI

---

individual's score across short periods of time, and the consistency with which the items measure the same dimension.

Andrews and Bonta report interrater reliability scores for the LSI-R ranging from  $r=.87$  to  $.94$  when the ratings took place within two months or less.<sup>50</sup> Test-retest reliability ranged from  $r=.95$  to  $.99$  when the instrument was administered by the same rater twice in under a month.<sup>51</sup> For internal consistency, the overall alpha value ranged from  $.64$  to  $.94$  with an average of  $.84$  across 13 studies.<sup>52</sup> The alpha coefficients for each subcomponent varied considerably as measured in the 13 studies, ranging from an average of  $.43$  to  $.78$ .

For the LS/CMI, Andrews, Bonta, and Wormith cite a combined interrater and test-retest reliability of  $r=.88$  for an average interval of 26 days between ratings.<sup>53</sup> For internal consistency, the overall alpha value ranged from  $.86$  to  $.92$  with an average of  $.89$  across eight studies.<sup>54</sup> As with the LSI-R, the alpha coefficients for each subcomponent varied considerably as measured across ten studies, ranging from an average of  $.44$  to  $.80$ .

**Potential for bias: gender.** Some researchers argue that some females follow gender-specific pathways to crime and that the gender-neutral LSI-R, developed on samples of primarily male offenders, has poor predictive validity for those types of females.<sup>55</sup> Reisig and his colleagues, for example, report that the LSI-R predicted recidivism for “economically motivated” female offenders (those similar to male offenders) but not for those who followed gendered pathways to crime in a sample of

women under community supervision in Minnesota and Oregon.<sup>56</sup>

However, the developers of the LSI-R claim that the tool is as reliable and as accurate in the prediction of reoffending for females as with males. They hold that the LS instruments were developed based on a general personality and cognitive social learning perspective of criminality and include separate norms for interpreting male and female total scores.<sup>57</sup> In addition, they cite several evaluations demonstrating the instruments' comparability in predicting male and female recidivism.<sup>58</sup> For example, a published, independent meta-analysis of 25 studies on a total of 14,737 female offenders did not uncover evidence of systematic gender bias in the predictive validity of the LSI-R, showing an average  $r = .35$  for women across these studies.<sup>59</sup> Sixteen of the 25 studies permitted a comparison of the LSI-R's predictive validity by gender; results of this analysis showed that the tool performed comparably for women ( $r_s = .27-.28$ ) and men ( $r_s = .24-.26$ ).

Van Voorhis and her colleagues found that the gender-neutral LSI-R assessment was strongly associated with new arrests in two samples of female probationers in Maui (AUC =  $.72$ ) and Minnesota (AUC =  $.71$ ).<sup>60</sup> However, they also noted that predictive validity increased when the LSI-R was supplemented with gender-responsive factors (AUC =  $.74$  for both sites). The specific factors adding to the improved validity differed somewhat for the two jurisdictions. The authors also found that some of the gender-responsive factors were more related to recidivism than the LSI-R factors, suggesting that treatment priorities

---

## Appendix: RNA Instrument Profile for the LSI-R and LS/CMI

---

for females might differ if using the gender-responsive supplement.

Andrews and his colleagues explored the predictive validity of each of the eight LS/CMI factors across five data samples and found each factor predictive of both male and female recidivism.<sup>61</sup> The only significant difference was the enhanced predictive validity of the substance abuse factor for females (AUC = .77 for females and .61 for males). However, they did find that the recidivism rates of low-risk females were substantially lower than those of low-risk men, prompting them to call for an exploration of different cut-off scores that would increase the number of women and decrease the number of men in low risk categories.

The LS test developers note that several gender-informed factors related to education/employment, family/marital (e.g., family conflict), and substance abuse already were in the LSI-R and were carried over to the LS/CMI.<sup>62</sup> In addition, the LS/CMI includes gender-informed items in Section 4: Other Client Issues and Section 5: Special Responsivity Considerations to assist in the development of effective case management plans. In their meta-analysis, Olver and his colleagues found that the LS total scores predicted general recidivism about equally well for men ( $r_w = .30$  and  $.30$ ) and women ( $r_w = .35$  and  $.31$ ). However, they also found that men tended to score higher on areas concerning “antisocial peers, lack of prosocial leisure activities, and substance abuse concerns linked to crime,” whereas women tended to score higher on areas concerning “more serious personal/emotional concerns, financial

problems, and family/marital difficulties” and faced “greater accommodation and education/employment concerns.”<sup>63</sup>

Authors encouraged careful consideration of possible gendered pathways to crime as part of a thorough case planning and program development process.

In sum, there is evidence demonstrating the predictive validity of the LS instruments for female offenders in general. However, variation in LS performance across jurisdictions as well as for specific types of female offenders in addition to the potential utility of individual gender-specific factors as supplements indicate the importance for additional research and local validation of the instruments to ensure their effectiveness.

**Potential for bias: race.** The LSI-R *U. S. Norms Manual Supplement* reports that race/ethnicity had no effect on the total LSI-R scores of community offenders (both male and female) and had a significant, though small, main effect (1% -2% of variability in scores) for inmates.<sup>64</sup> All of the analyses (male, female, community offenders, and inmates) compared Caucasian and African Americans except for male inmates which also included Hispanic, Asian, and Native American offenders. Olver and his colleagues reported significantly smaller effect sizes among ethnic minority offenders ( $r_w = .23$  and  $.23$ ) than among non-minorities ( $r_w = .32$  and  $.29$ ), but concluded that these differences were too small in magnitude to be substantively meaningful.<sup>65</sup>

The findings of additional studies vary. For example, in a study involving 445 African American and Hispanic male inmates released into halfway houses in New Jersey, the predictive validity for rearrest within two

---

## Appendix: RNA Instrument Profile for the LSI-R and LS/CMI

---

years was  $r = .08$  for African Americans and  $r = .02$  for Hispanic offenders.<sup>66</sup> The predictive validity for reconviction within two years was  $r = .11$  for African American offenders and  $r = .04$  for Hispanic offenders. The researchers found these correlations low compared to other published studies and concluded that “further analysis of the use of the LSI-R on minority offender populations is warranted and encouraged.”<sup>67</sup>

Another study examined the predictive validity of the LSI-R for a sample of 696 male offenders (72% African American, 15% Hispanic, 13% Caucasian) released from prison in New Jersey.<sup>68</sup> The outcome measure was rearrest within one year. Predictive validity was best for African American offenders (AUC = .61) followed by Caucasian offenders (AUC = .55) and then Hispanic offenders (AUC = .54). The researchers found that African Americans were more likely to be overclassified (rearrest predicted but did not occur), and Hispanics and Caucasians were more likely to be underclassified (no rearrest predicted but rearrest occurred). An additional study reported an overall trend toward more overclassification *and* underclassification for African Americans in a sample of 532 male residents at a federal community corrections center.<sup>69</sup> The sample was 52% African American, 33% Caucasian, and 12% Hispanic. The extent of classification errors varied by the cutoff score and performance measure (i.e., program success or disciplinary incidents) used. The author noted that the low base rate for program failure (11%) and potential reliability issues in scoring the LSI-R may have influenced the results. He concluded that the results highlighted the

need for correctional facilities to validate the instrument on their own populations.

Another study examined the predictive validity of the LSI-R for Native Americans.<sup>70</sup> The study followed 403 community-supervised offenders (56% White and 35% Native American) in the northern midwest for 17 months. The researchers reported predictive validity values of  $r = .18$  for all the offenders,  $r = .23$  for the White offenders, and  $r = .11$  for Native American offenders. Predictive validity was lowest for Native American females with an  $r = -.13$ ; predictive validity for male Native Americans was  $r = .19$ . The researchers suggested additional research to determine whether there are (a) more relevant factors for predicting antisocial behavior among Native Americans, (b) different results when stronger outcome measures (e.g., reconviction rather than rearrest) are used, and (c) different results with a larger sample of Native American women than the current sample of 40. They also questioned whether differences in responsivity factors among Native Americans and the race/ethnicity of the professionals conducting the assessments might affect assessment results.

As with gender, these studies highlight the need for additional research on the predictive validity of the LS instruments for various racial and ethnic groups as well as the importance of validating the instrument for use in specific settings.

***Independent validation.*** The LSI-R has been independently validated across multiple studies and jurisdictions as noted under the “predictive validity” section above. A study investigating the variability in the magnitude of predictive validity estimates

---

## Appendix: RNA Instrument Profile for the LSI-R and LS/CMI

---

found that larger estimates are associated with studies involving LS authors, those conducted in Canada, and those with longer follow-up periods.<sup>71</sup> The authors suggest that the findings are due, in part, to the integrity with which the instruments are used. This explanation was supported in a study by Flores and his colleagues who found that the predictive validity of the LSI-R increased with formal staff training and agency experience with the tool.<sup>72</sup>

### PRACTICAL CONSIDERATIONS.

---

**Vendor and instrument cost.** The LS tools are available for purchase from Multi-Health Systems (MHS).<sup>73</sup>

**Menu of other services.** MHS offers a wide array of services, training, and technical assistance to support the use of LS instruments.

- **IT SERVICES.** Software is available through MHS for completing and scoring the LS instruments. The software can be purchased on a per-use basis, site-licensed, or customized to fit with a jurisdiction's existing database. For larger, jurisdiction-wide implementation, MHS recommends using the Software Developer's Kit (SDK) to integrate the LS tool into the jurisdiction's case management system.<sup>74</sup> This option allows the data to be stored in-house and accessed at any time and requires no maintenance, administration, or technical assistance fees.<sup>75</sup>
- **TECHNICAL ASSISTANCE.**<sup>76</sup> For jurisdictions that opt to use the SDK, MHS has a team of programmers available to help local programmers

incorporate the LS instrument into their case management systems. Toll free assistance is available for those using the standard software package (Smartlink) as well.

Additional assistance is available from a team of researchers to answer questions about the psychometric properties of the instruments. MHS also maintains a Community of Users listserv which serves as a forum to ask questions and share information on policies, procedures and practices.<sup>77</sup> LS users also can submit questions about the tools to the instrument developers.

- **VALIDATION SERVICES.** MHS will norm the LS instruments on the local population once 1,000 assessments have been conducted. There is no additional cost for this service.<sup>78</sup>
- **USER TRAINING.** MHS maintains a training network of certified LS trainers who offer employee training and train-the-trainer programs. The latter saves a jurisdiction the cost of bringing in an outside trainer for each new employee and booster training program. Training costs vary. The jurisdiction submits a request to the network, and trainers bid on the request given the jurisdiction's budget, timing, and needs.<sup>79</sup>

**User qualifications.** To be qualified, test administrators must be trained by an MHS-approved trainer or training program unless they have completed graduate level courses in tests/measurement or can document similar training. Test administrators who do not meet the qualifications must be supervised by a qualified administrator.<sup>80</sup>

## Appendix: RNA Instrument Profile for the LSI-R and LS/CMI

---

**Administration time.** The MHS website lists the administration time as 30-45 minutes for the LSI-R and 20-30 minutes for the LS/CMI.<sup>81</sup> The test developers estimate that the client interview can take an hour to an hour and a half.<sup>82</sup>

**Modes of administration.** Information for the LS instruments is collected through a structured interview with the offender, reviews of files and official records, interviews with collaterals such as family members, and, if available, psychological test data.<sup>83</sup>

**Quality assurance.** Several quality assurance considerations follow.

- **OVERRIDE POLICY.** The LS instruments allow professionals to override the quantitative assessment if they identify factors they think deserve special consideration in determining the offender's risk level.<sup>84</sup> If the override option is used, the test administrator is required to provide a written explanation for changing the initial score. The LS/CMI manual notes that overrides are expected in fewer than 10% of cases.<sup>85</sup> The manual also indicates that aggravating and mitigating factors identified in other sections of the LS/CMI may be used to inform and justify the override decision.<sup>86</sup> However, additional research by Wormith and his colleagues indicates that predictive validity decreased when the override was used.<sup>87</sup> The authors found that test administrators used the override much more frequently to increase an offender's risk level than to decrease it and cautioned against the overuse of the practice.
- **FIDELITY.** Because LS instruments require administrator expertise to properly score, formal training is critical to the effective implementation of the instrument. The test developers recommend that initial training be supplemented with periodic booster sessions and audit checks of test administrators' assessments.<sup>88</sup> Agencies with automated databases also can look for systematic trends (e.g., frequent use of overrides for certain types of offenders) in scoring that suggest the need for consultation and/or additional training.<sup>89</sup> Agency staff also uses the MHS Community of Users listserv to discuss quality assurance issues and share strategies to monitor quality.<sup>90</sup>
- **INSTRUMENT REVALIDATION.** The test developers do not have recommendations for the frequency with which LS instruments should be revalidated for a jurisdiction, noting that it depends on an agency's workload and resources. They revalidate the instrument in Ontario approximately every five years and suggest that as a general rule of thumb.<sup>91</sup>

---

### ENDNOTES

<sup>1</sup> See p. 47 in Andrews, D. A., & Bonta, J. (2006). *The psychology of criminal conduct, (4th ed.)*. Cincinnati: Anderson.

<sup>2</sup> See p. 47 in Andrews & Bonta (2006) at endnote 1.

<sup>3</sup> See pp. 22-23 and 151 in Andrews, D. A., Bonta, J., & Wormith, J. S. (2004). *The Level of Service/Case Management Inventory (LS/CMI): User's manual*. Toronto: Multi-Health Systems

## Appendix: RNA Instrument Profile for the LSI-R and LS/CMI

---

<sup>4</sup> See Andrews & Bonta (2006) at endnote 1, p. 48.

<sup>5</sup> See Andrews & Bonta (2006) at endnote 1, p. 48.

<sup>6</sup> See Andrews & Bonta (2006) at endnote 1, p. 283.

<sup>7</sup> See Andrews, Bonta, & Wormith (2004) at endnote 3, p. 155.

<sup>8</sup> See Andrews & Bonta (2006) at endnote 1, p. 48.

<sup>9</sup> See Andrews & Bonta (2006) at endnote 1, p. 48.

<sup>10</sup> See Andrews, Bonta, & Wormith (2004) at endnote 3, p. 11.

<sup>11</sup> S. Wormith, personal communication, February 8, 2014.

<sup>12</sup> Andrews, D. A. (1982). *The Level of Supervision Inventory (LSI): The first follow-up*. Toronto: Ontario Ministry of Correctional Services.

<sup>13</sup> Andrews, D. A. & Bonta, J. (1995). *The Level of Service Inventory—Revised: User’s manual*. Toronto: Multi-Health Systems.

<sup>14</sup> See Andrews, Bonta, & Wormith (2004) at endnote 3.

<sup>15</sup> Andrews, D. A., Bonta, J., & Wormith, J. S. (2010). The Level of Service (LS) assessment of adults and older adolescents. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of Violence risk assessment* (pp. 199-225). New York: Taylor and Francis Group.

<sup>16</sup> See Andrews & Bonta (1995) at endnote 13, p. viii and Andrews, Bonta, & Wormith (2004) at endnote 3, p. 7.

<sup>17</sup> See Andrews, Bonta, & Wormith (2010) at endnote 15.

<sup>18</sup> See Andrews, Bonta, & Wormith (2004) at endnote 3, p. 1.

<sup>19</sup> See Andrews & Bonta (1995) at endnote 13, p. 1 and Andrews, Bonta, & Wormith (2004) at endnote 3, p. 1.

<sup>20</sup> J. S. Wormith (personal communication, September 19, 2012).

<sup>21</sup> See Andrews (1982) at endnote 12, p. 2 and Andrews, Bonta, & Wormith (2010) at endnote 15, p. 205.

<sup>22</sup> See Andrews (1982) at endnote 12.

<sup>23</sup> Subsequent factor analyses conducted on the LSI-R produced inconsistent results regarding the instrument’s underlying constructs. Various studies have yielded one, two, and three-factor results. All included a factor related to the propensity to engage in crime. See Andrews, Bonta, & Wormith (2010) at endnote 15, p. 211.

<sup>24</sup> See Andrews & Bonta (1995) at endnote 13, p. 13.

<sup>25</sup> See p. 3 in Andrews, D. A., & Bonta, J. L. (2003). *LSI-R: U. S. norms manual supplement*. North Tonawanda, NY: Multi-Health Systems.

<sup>26</sup> See Andrews & Bonta (2003) at endnote 25, p. 3.

<sup>27</sup> See Andrews, Bonta, & Wormith (2010) at endnote 15, p. 205 and Andrews, Bonta, & Wormith (2004) at endnote 3, p. 2.

<sup>28</sup> See Andrews, Bonta, & Wormith (2004) at endnote 3, p. 3.

<sup>29</sup> See Andrews, Bonta, & Wormith (2004) at endnote 3, p. xiv. Information also provided by J. Stephen Wormith (personal communication, September 19, 2012).

<sup>30</sup> See Andrews, Bonta, & Wormith (2004) at endnote 3, p. xiv.

<sup>31</sup> The LS/CMI manual (see endnote 3) also reports North American norms. However, the test developers no longer recommend using those norms. The numerous differences observed between the Canadian and U. S. samples make it difficult to interpret a score based on the pooled samples. S. Wormith, personal communication, May 15, 2014.

<sup>32</sup> See Andrews, Bonta, & Wormith (2004) at endnote 3, pp. 169, 175. The differences in the n sizes across tables is due to cases in which only total scores or a designated risk level without item data were or were not included. J. S. Wormith, personal communication, May 15, 2004.

<sup>33</sup> See Andrews, Bonta, & Wormith (2004) at endnote 3, p.3.

<sup>34</sup> See Andrews, Bonta, & Wormith (2004) at endnote 3, p. 3.

## Appendix: RNA Instrument Profile for the LSI-R and LS/CMI

---

<sup>35</sup> See Andrews & Bonta (2003) at endnote 25, p. 4. The cutoff scores differ for institutional classification.

<sup>36</sup> See Andrews, Bonta, & Wormith (2004) at endnote 3, p. 36.

<sup>37</sup> See Andrews & Bonta (1995) at endnote 13, p. 14.

<sup>38</sup> See Andrews, Bonta, & Wormith (2004) at endnote 3, p. 36.

<sup>39</sup> The authors also reported significant differences in effect size magnitude by geographic region, which was highest in Canadian samples ( $r_w = .38$  and  $.43$  for fixed- and random-effects models, respectively), followed by non-North American samples ( $r_w = .30$  and  $.29$ ), and United States samples ( $r_w = .20$  and  $.22$ ). Olver, M., Stockdale, K., & Wormith, J. (2014). Thirty years of research on the Level of Service scales: A meta-analytic examination of predictive accuracy and sources of variability. *Psychological Assessment, 26*, 156-176.

<sup>40</sup> Vose, B., Cullen, F. T., & Smith, P. (2008). The empirical status of the Level of Service Inventory. *Federal Probation, 72* (3), 22-29.

<sup>41</sup> See pp. 24-25 in Vose, Cullen, & Smith (2008) at endnote 40. The range reported excludes studies examining program completion and parole violations.

<sup>42</sup> Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology, 34*, 575-607. The effect size adjusted for sample size is  $r = .33$ .

<sup>43</sup> Gendreau, P., Goggin, C., & Smith, P. (2002). Is the PCL-R really the “unparalleled” measure of offender risk? A lesson in knowledge cumulation. *Criminal Justice and Behavior, 29*, 397-426. The effect size adjusted for sample size is  $.39$ . The study also provided a mean effect size of  $r = .26$  ( $r = .28$ , adjusted for sample size) for predicting violent recidivism. A subsequent meta-analysis by Campbell and her colleagues reports a mean effect size for violent recidivism of  $r = .25$  ( $r = .28$ , adjusted for sample size). See Campbell, M. A., French, S.,

& Gendreau (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior, 36*, 567-590.

<sup>44</sup> Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime and Delinquency, 52*, 7-27. The authors also report a mean effect size for violent recidivism of  $r = .25$ .

<sup>45</sup> For the federal probation study, see Flores, A. W., Lowenkamp, C. T., Smith, P., & Latessa, E. J. (2006). Validating the Level of Service Inventory—Revised on a sample of federal probationers. *Federal Probation, 70*(2), 44-49. For the Iowa probation and parole study, see Lowenkamp, C. T., & Bechtel, K. (2007). The predictive validity of the LSI-R on a sample of offenders drawn from the records of the Iowa Department of Corrections data management system. *Federal Probation, 71*(3), 25-29.

<sup>46</sup> See Andrews, Bonta, & Wormith (2010) at endnote 15, p. 213. This profile of the LS/CMI is focused on Section 1 “General Risk/Need Factors.” The LS/CMI manual (see pp. 122-123 in Andrews, Bonta, & Wormith, 2004, at endnote 3) also provides limited information about the predictive validity of Sections 2 through 5.

<sup>47</sup> Girard, L., & Wormith, J. S. (2004). The predictive validity of the Level of Service Inventory—Ontario Revision on general and violent recidivism among various offender groups. *Criminal Justice and Behavior, 31*, 150-181. The effect size for violent conviction is  $r = .28$ .

<sup>48</sup> See Andrews, Bonta, & Wormith (2010) at endnote 15, pp. 213-214.

<sup>49</sup> Arnold, T. (2007). *Dynamic changes in the Level of Service Inventory-Revised (LSI-R) scores and the effects on prediction accuracy*. Master’s thesis, St. Cloud University, St. Cloud, MN.

<sup>50</sup> See Andrews & Bonta (1995) at endnote 13, p. 35.

## Appendix: RNA Instrument Profile for the LSI-R and LS/CMI

---

- <sup>51</sup> See Andrews & Bonta (1995) at endnote 13, p. 35.
- <sup>52</sup> See Andrews, Bonta, & Wormith (2010) at endnote 15, pp. 206-208.
- <sup>53</sup> See Andrews, Bonta, & Wormith (2004) at endnote 3, pp. 114-115.
- <sup>54</sup> See Andrews, Bonta, & Wormith (2010) at endnote 15, pp. 206-208.
- <sup>55</sup> See, for example, Holtfreter, K. & Cupp, R. (2007). Gender and risk assessment: The empirical status of the LSI-R for women. *Journal of Contemporary Criminal Justice*, 23, 363-382; Reisig, M. D., Holtfreter, K., & Morash, M. (2006). Assessing recidivism risk across female pathways to crime. *Justice Quarterly*, 23, 384-405.
- <sup>56</sup> See Reisig, et al. (2006) at endnote 55, p. 397. The predictive validity was  $r = .24$  for economically motivated offenders and  $r = -.13$  for gendered pathway offenders.
- <sup>57</sup> See Andrews & Bonta (1995) at endnote 13, p. 48 and Andrews, Bonta, & Wormith (2004) at endnote 3, p. 143.
- <sup>58</sup> Andrews, D. A., Bonta, J., & Wormith, J. S. (2009). *LS/CMI: A gender-informed risk/need/responsivity assessment*. North Tonawanda, NY: Multi-Health Systems.
- <sup>59</sup> Smith, P., Cullen, F. T., & Latessa, E. J. (2009). Can 14,737 women be wrong? A meta-analysis of the LSI-R and recidivism for female offenders. *Criminology & Public Policy*, 8, 183-208.
- <sup>60</sup> Van Voorhis, P., Wright, E. M., Salisbury, E., & Bauman, A. (2010). Women's risk factors and their contributions to existing risk/needs assessment: The current status of a gender-responsive supplement. *Criminal Justice and Behavior*, 37, 261-288.
- <sup>61</sup> Andrews, D. A., Guzzo, L., Raynor, P., Rowe, R. C., Rettinger, L. J., Brews, A., & Wormith, J. S. (2012). Are the major risk/need factors predictive of both female and male reoffending? A test with the eight domains of the Level of Service/Case Management Inventory. *International Journal of Offender Therapy and Comparative Criminology*, 56(1), 113-133.
- <sup>62</sup> Andrews, D. A., Bonta, J., & Wormith, J. S. (2009). *Level of Service/Case Management Inventory (LS/CMI). Supplement: A gender-informed risk/need/responsivity assessment*. North Tonawanda, NY: Multi-Health Systems.
- <sup>63</sup> Olver et al. (2013) at endnote 39, p. 14.
- <sup>64</sup> See Andrews & Bonta (2003) at endnote 25, p. 10.
- <sup>65</sup> Olver et al. (2013) at endnote 39.
- <sup>66</sup> Schlager, M. D., & Simourd, D. J. (2007). Validity of the Level of Service Inventory—Revised (LSI-R) among African American and Hispanic male offenders. *Criminal Justice and Behavior*, 34, 545-554.
- <sup>67</sup> See Schlager & Simourd (2007) at endnote 66, p. 553.
- <sup>68</sup> Fass, T. L., Heilbrun, K., Dematteo, D., & Fretz, R. (2008). The LSI-R and the COMPAS: Validation data on two risk-needs tools. *Criminal Justice and Behavior*, 35, 1095-1108.
- <sup>69</sup> Whiteacre, K.W. (2006). Testing the Level of Service Inventory-Revised (LSI-R) for racial/ethnic bias. *Criminal Justice Policy Review*, 17, 330-342.
- <sup>70</sup> Holsinger, A. M., Lowenkamp, C. T., & Latessa, E. J. (2006). Exploring the validity of the Level of Service Inventory-Revised with Native American offenders. *Journal of Criminal Justice*, 34, 331-337.
- <sup>71</sup> Andrews, D. A., Bonta, J., Wormith, J. S., Guzzo, L., Brews, A., Rettinger, J., & Rowe, R. (2011). Sources of variability in estimates of predictive validity: A specification with Level of Service general risk/need. *Criminal Justice and Behavior*, 38, 413-432.
- <sup>72</sup> Flores, A. W., Lowenkamp, C. T., Holsinger, A. M., & Latessa, E. J. (2006). Predicting outcome with the Level of Service Inventory-Revised: The importance of implementation integrity. *Journal of Criminal Justice*, 34, 523-529. Also see Austin, J., Coleman, D., Peyton, J., & Johnson, K. D. (2003). *Reliability and validity study of the LSI-R risk assessment instrument*. Washington, DC: The Institute on Crime,

## Appendix: RNA Instrument Profile for the LSI-R and LS/CMI

---

Justice and Corrections at the George Washington University. Austin and his colleagues found that reliability increased after additional training was provided to the test administrators.

<sup>73</sup> Information on the LSI-R is available at <http://www.mhs.com/product.aspx?gr=saf&prod=lsi-r&id=overview>; and information on the LS/CMI is available at <http://www.mhs.com/product.aspx?gr=saf&prod=ls-cmi&id=overview>. According to an undated report provided by Tammy Howell of MHS, substantial discount rates are available for states and counties interested in adopting LS instruments. See, Howell, T. (n.d.). *LSI-R, LS/RNR, and LS/CMI documentation*. North Tonawanda, NY: Multi-Health Systems. Retrieved from <http://www.scstatehouse.gov/archives/citizeninterestpage/SentencingReformCommission/SentencingReform.php>

<sup>74</sup> See Andrews, Bonta, & Wormith (2010) at endnote 15, p. p. 200.

<sup>75</sup> See Howell (n.d.) at endnote 73: MHS charges a one-time fee to purchase the SDK and charges an annual fee based on the number of assessments conducted. The per assessment fee is based on a sliding scale according to the number of assessments purchased. A standard software package (Smartlink) also is available.

<sup>76</sup> Information for this section comes from Howell (n.d.) at endnote 73.

<sup>77</sup> The Community of Users also was discussed by the test developers: J. Bonta and S. Wormith, personal communication, April 17, 2012.

<sup>78</sup> J. Bonta and S. Wormith, personal communication, April 17, 2012. Also see Howell (n.d.) at endnote 73.

<sup>79</sup> See Howell (n.d.) at endnote 73 and the MHS web site at [https://ecom.mhs.com/\(S\(go2ocy45brgsoq55mnrk5m55\)\)/saf\\_om.aspx?id=Training](https://ecom.mhs.com/(S(go2ocy45brgsoq55mnrk5m55))/saf_om.aspx?id=Training).

<sup>80</sup> See Andrews, Bonta, & Wormith (2004) at endnote 3, p. 6.

<sup>81</sup> See MHS website at [https://ecom.mhs.com/\(S\(oazaozbybrdg5uz5fc1yzj55\)\)/product.aspx?gr=saf&prod=lsi-r&id=overview](https://ecom.mhs.com/(S(oazaozbybrdg5uz5fc1yzj55))/product.aspx?gr=saf&prod=lsi-r&id=overview) and [https://ecom.mhs.com/\(S\(2s1oyojq5lrrttayrndkiko45\)\)/product.aspx?gr=saf&prod=ls-cmi&id=overview](https://ecom.mhs.com/(S(2s1oyojq5lrrttayrndkiko45))/product.aspx?gr=saf&prod=ls-cmi&id=overview).

<sup>82</sup> J. Bonta and S. Wormith, personal communication, April 17, 2012.

<sup>83</sup> See Andrews, Bonta, & Wormith (2010) at endnote 15, p.200.

<sup>84</sup> See Andrews & Bonta (1995) at endnote 13, p. 12 and Andrews, Bonta, & Wormith (2004) at endnote 3, p. 4.

<sup>85</sup> See Andrews, Bonta, & Wormith (2004) at endnote 3, p. 4.

<sup>86</sup> See Andrews, Bonta, & Wormith (2004) at endnote 3, p. 123.

<sup>87</sup> Wormith, J. S., Hogg, S., & Guzzo, L. (2012). The predictive validity of a general risk/needs assessment inventory on sexual offender recidivism and an exploration of the professional override. *Criminal Justice and Behavior*, 39, 1511-1538. The authors found that predictive validity decreased for all offenders but was particularly lowered for sexual offenders.

<sup>88</sup> See Andrews, Bonta, & Wormith (2010) at endnote 15, pp. 209-210.

<sup>89</sup> J. Bonta and S. Wormith, personal communication, April 17, 2012.

<sup>90</sup> J. Bonta and S. Wormith, personal communication, April 17, 2012.

<sup>91</sup> J. Bonta and S. Wormith, personal communication, April 17, 2012.

# The Offender Screening Tool (OST)

---

## OST GLOSSARY OF TERMS

---

<b>Risk</b>	Authors adhere to Andrews and Bonta’s risk principle, stating that "supervision strategies should prioritize treatment and probation resources for higher risk offenders." <sup>1</sup>
<b>Static risk</b>	Authors use this term to describe risk factors that "contribute to an individual's risk to reoffend but cannot be changed." <sup>2</sup>
<b>Dynamic risk</b>	In combination with static (historical) risk factors, dynamic (changeable) risk factors have been found to be significant predictors of recidivism. The authors state that dynamic factors help "identify potential targets for treatment" and "contribute to an individual's overall risk to reoffend." <sup>3</sup>
<b>Needs</b>	Authors adhere to Andrews and Bonta’s needs principle, stating that "probation strategies should target interventions to criminogenic needs. Supervision should address the offenders’ needs that are directly linked to criminal behavior." <sup>4</sup>
<b>Responsivity</b>	Authors acknowledge Andrews and Bonta’s responsivity principle, stating that "probation staff should be responsive to temperament, learning style, motivation, culture, and gender when assigning programs". <sup>5</sup>
<b>Protective factors</b>	Term not used.
<b>Strengths</b>	Term not used.
<b>Recidivism</b>	In the independent assessment of OST <sup>6</sup> , the evaluators used five separate measures of recidivism: (1) petition to revoke, (2) petition to revoke with new arrest, (3) revoked, (4) any arrest, and (5) felony arrest.

---

## HISTORY & CURRENT USE.

---

**Creation.** In 1996, the Maricopa County Adult Probation Department (MCAPD) in Arizona reviewed existing offender assessment practices as part of its commitment to research-based practices.<sup>7</sup> This review prompted the MCAPD to seek a more meaningful offender risk and needs assessment tool. In response, the staff of MCAPD, with the assistance of research

consultant Dr. David Simourd, developed the Offender Screening Tool (OST) in 1998.

**Current use.** Although developed for use in Maricopa County, the OST was subsequently validated on the probation population statewide and was adopted by the Arizona Administrative Office of the Courts (AOC) for statewide use with probationers in January, 2005.<sup>8</sup> Prior to the OST, most counties in the state were using variants of

## Appendix: RNA Instrument Profile for the OST

---

the Wisconsin risk and needs assessment tool, which had never been validated for the Arizona probation population.<sup>9</sup> In addition, there was evidence that probation officers across the state were not using the instruments consistently, nor were the results being used to inform decisions about the level of services to be received.<sup>10</sup>

The OST is also used in local probation departments in Virginia with misdemeanor offenders.<sup>11</sup>

### DEVELOPMENT.

---

***Instrument purpose.*** MCAPD sought an instrument that would assess both risk and needs of the offender using static and dynamic measures directly related to the key predictors of criminal behavior. The goal was to implement an instrument that would gauge the likelihood of individual reoffending and also identify specific offender needs that could be used to inform more effective treatment and service delivery. Additionally, to increase the likelihood that the tool would be used consistently and as intended, MCAPD wanted a tool that probation staff viewed as meaningful.<sup>12</sup>

Rather than draw on a pre-existing risk and needs assessment instrument, MCAPD decided to create its own tool. This decision reflected several factors, including a concern about the annual cost of using an existing proprietary tool given the large number of assessments done each year, the need to identify a tool that was valid for use with the local population of offenders, and a strong desire to involve probation staff in the development of the tool. At that time,

MCAPD had also decided to reengineer the operation of their presentence division. Through reengineering, the presentence process was streamlined, duplicated effort was eliminated, and the OST system was introduced to the department.

With the OST system, probation officers make use of three main assessment tools: a full assessment, a reassessment, and a brief screener. The OST, the full assessment tool, is administered at the presentence phase to identify offender behaviors over the previous 12 to 36 months. Results are used to guide case management decisions. To capture the effect of probationary intervention and inform case management decisions over time, Arizona employs the Field Reassessment Offender Screening Tool (FROST), nearly identical to the OST in items and scoring, to reassess offenders for changes in risk and needs over time.<sup>13</sup> The FROST is designed to be conducted at 6 month intervals. Completing either the full assessment or the reassessment requires a review of the case file and an interview with the offender. Some judgment is needed to score items on the instrument.

An abbreviated version of the OST, the Modified Offender Screening Tool (MOST), was developed for expedited use and draws on 8 items from the OST.<sup>14</sup> Designed as a relatively quick screening tool, higher scores on the MOST are a signal to probation officers to administer the full OST.

***Approach to instrument development.*** In creating the OST, the developers used an approach that was more theoretically than statistically driven.<sup>15</sup> From this framework, they incorporated factors related to both risk

## Appendix: RNA Instrument Profile for the OST

---

and needs in a single instrument that they believed reflected the latest thinking about the psychology of criminal conduct. Related in design to the Level of Service Inventory-Revised (LSI-R), OST employs a similar class of variables drawn from correctional and developmental literature and from existing meta-analytic research identifying the strongest predictors of recidivism.<sup>16</sup>

Assessment questions were based on these variables identified in the research literature as related to criminal behavior.<sup>17</sup> Although the developers assert that all factors on the OST are related to recidivism, those categories that are stronger predictors are given more weight (i.e., more items) and therefore have greater influence on the overall risk score.<sup>18</sup>

In addition to this theoretical relationship, instrument developers sought items that showed a statistically significant relationship with recidivism, had face validity to facilitate buy-in from the court community, and that could be easily scored by probation officers to ensure consistent and proper use of the tool. They also sought to include items that were relevant in the treatment process and strongly preferred dynamic over static items. The final OST is comprised of items that are 61% dynamic.<sup>19</sup>

### CONTENT.

---

**Structure.** The OST and FROST each produce a single overall score from a set of nine subscales. This overall score is used to determine the offender's recidivism risk level. Overall scores are positively related to multiple measures of the offender's risk of recidivism, with the two primary outcome

measures being (1) petition to revoke and (2) any new arrest.<sup>20</sup>

When using the full instrument, probation officers are told that the primary needs areas (those that require intervention through case planning) are those identified by the nine subscales. Scores for each of the nine subscales are used to identify and prioritize the offender's needs for case planning and service provision. A 10th section of the OST and FROST contains two additional items, referred to as responsivity factors. They are not criminogenic and are not incorporated into the computation of overall risk or individual needs.<sup>21</sup>

**Items and domains.** The OST is comprised of 42 items across the nine different risk and need subscales (or domains). Each domain is comprised of 2-9 items that may be static or dynamic. The nine domains include: vocational/financial (5 items), education (3 items), family & social relationships (8 items), residence & neighborhood (2 items), alcohol (3 items), drug abuse (3 items), mental health (2 items), attitude (7 items), and criminal behavior (9 items).

The tenth section on responsivity factors includes two additional physical/medical health items (for a total of 44 items) and is used to identify whether or not health-related concerns may pose potential barriers to successful offender treatment.

In general, OST items are scored on the basis of patterns of behavior rather than a single incident (e.g., a single incident of alcohol use should not necessarily be coded as problematic use). Probation officers are encouraged to have at least one or two

## Appendix: RNA Instrument Profile for the OST

---

reasons that explain why each OST item is scored as it is for the offender.

**Reporting and cutoffs.** When first introduced, the OST categorized offenders into one of three levels of risk (low, moderate, high) based on overall scores. The cutoff values used to create these three categories were estimated based on the cutoff scores used in the LSI-R. Following an initial period of use, these cutoffs were revised based on actual OST data from the local probation population in Maricopa County.<sup>22</sup>

When the OST was adopted statewide, the cutoff scores were reexamined. Results from a 2008 independent statewide validation study indicated that the range of scores in the moderate risk category was too large to sufficiently differentiate offenders.<sup>23</sup> Based on these findings, the OST risk categories were again revised, this time expanding from three risk categories to four. In addition, separate cutoff values were established for men and women. These new cutoff values were as follows: For males, low (1-5 points), moderate (6-10 points), moderate-high (11-17 points), and high (18+ points); for females, low (0-8 points), moderate (9-13 points), moderate-high (14-20 points), and high (21+ points). The low-risk cutoff values were selected to align with a 15 percent failure (recidivism) rate.

Unlike some tools, the OST does not produce similar ranking categories to identify level of need in each domain. Rather, probation officers are encouraged to target needs identified by dynamic items in high-scoring domains.

### INSTRUMENT RELIABILITY AND VALIDITY.

---

**Populations studied.** Following the creation of OST on a construction sample of male and female Maricopa County probationers,<sup>24</sup> the instrument was validated in 2003 on a statewide sample of male and female probationers<sup>25</sup> and independently validated on another Arizona statewide sample in 2008.<sup>26</sup> A statewide validation study has also been completed in Virginia.<sup>27</sup> In these validation studies, researchers selected representative samples of offenders who had a case closed within a suitable timeframe to allow for an evaluation of probation outcome (e.g., at least six months). The validation study in Virginia, for example, distinguished offenders based on age, sex, race/ethnicity, criminal history, current charge, and geographic location.

**Predictive validity.** The developers found that prior OST risk scores were significantly higher for offenders whose current probation status was deemed "unsatisfactory" vs. those whose current behavior was found to be "satisfactory."<sup>28</sup>

A more rigorous analysis of the OST's predictive validity was undertaken in 2008.<sup>29</sup> With respect to recidivism, researchers found that the OST works best in Arizona as a predictor of petitions to revoke ( $r=.23$ ) and less well as a predictor of any arrest ( $r=.12$ ).<sup>30</sup> In Virginia, one outcome variable was examined: probation closure type.<sup>31</sup> Closure type was coded as a) successful, b) transfer in-sent back, or c) unsuccessful. A linear relationship between OST scores and outcome was expected because greater OST scores are designed to be reflective of greater

## Appendix: RNA Instrument Profile for the OST

---

criminal need issues. A statistically significant relationship between the OST score and outcome was found ( $r=.19$ ).

**Reliability.** Lowenkamp and colleagues reported levels of inter-rater agreement in Arizona above 90% for 25 of the 42 OST items.<sup>32</sup> Lower percentages of agreement tended to emerge from items that required the assessor to count or identify times of occurrences (e.g., two or less times unemployed,) and for items that required more professional discretion (e.g., client being in denial about alcohol use). The evaluators recommended that rater consistency could be improved with more training.

**Potential for bias.** Simourd examined the differential validity of the OST on males and females and found no significant differences in overall scores, but significant differences within certain domains of the tool.<sup>33</sup> Males were found to have significantly greater scores on the Education, Alcohol, and Criminal Behavior domains, while females had significantly greater scores on Vocational/Financial, Family and Social Relationships, and Mental Health domains. He found no significant differences by county or by type of offense. Simourd concluded that the observed gender differences were small in practical terms and therefore made no recommendations for change.

Lowenkamp and colleagues examined the differential validity of OST on sex and ethnicity (Hispanic vs. non-Hispanic) and concluded that the tool performs adequately for all subgroups in predicting petitions to revoke, but less well for other measures of

recidivism.<sup>34</sup> They suggested altering cutoff scores to improve predictive validity but did not make any further recommendations. Following the evaluation, the OST moved to four categories of risk and established different cutoff values for men and women.

**Independent validation.** One independent validation has been conducted to date.<sup>35</sup>

### PRACTICAL CONSIDERATIONS.

---

**Vendor and instrument cost.** The OST system is non-proprietary. For more information, contact Dr. Jennifer Ferguson of MCPAD ([jferguso@apd.maricopa.gov](mailto:jferguso@apd.maricopa.gov)) or Dr. David Simourd of ACES Inc. ([dave@acesink.com](mailto:dave@acesink.com)).

**Menu of other services.** Not applicable for this non-proprietary tool, although independent consultants have offered research support and validation services.

**User qualifications.** The OST is administered by the Arizona Adult Probation Department (APD) presentence division. Individual probation officers administer the reassessment (FROST).<sup>36</sup>

Following the 2008 reliability and validation study,<sup>37</sup> APD instituted mandatory initial and refresher training requirements for presentence division staff and probation officers. All probation officers are trained on the instrument. Presentence screeners receive training on interviewing skills and, after completing several interviews in the field, participate in focus groups to exchange feedback and refine their OST administration skills. Probation officers must complete a three-year refresher training, which includes a review the OST system and

## Appendix: RNA Instrument Profile for the OST

---

addresses the topic of developing appropriate case plans.<sup>38</sup> In addition, the training program reviews the OST and FROST Scoring Guides, which provide descriptions of and scoring tips for all items.

**Administration time.** Developers say the OST and FROST take, on average, about 25 minutes to complete.<sup>39</sup>

**Modes of administration.** Information used to complete the OST and FROST is drawn from a structured interview that relies partly on offender self-report. The administering presentence screener or probation officer leads the interview. The computerized OST system automatically calculates assessment results.

**Quality assurance.** When adopting any offender assessment tool, jurisdictions must be prepared to ensure appropriate use and maintenance over time. Protocols established by Maricopa County and the state of Arizona Probation Departments are described below.

- **OVERRIDE POLICY.** The stated goal in Arizona is to minimize the number of overrides to the OST recommendations.<sup>40</sup> When first implemented, the developers indicated that an override of OST results should occur in no more than 10% of cases. Currently, there is no specific numerical target and no systematic effort to track overrides. The decision to override the instrument recommendation is made on a case-by-case basis when the probation officer believes it is justified.
- **FIDELITY.** Reliability in the use of the instrument depends to a great extent on

training. In Maricopa County, a refresher training system has been developed to improve scoring consistency among presentence screeners and probation officers. These users first view an educational refresher training video online and then complete a scoring test. If the user does not pass the scoring test, they are required to attend an in-person classroom refresher training course and retake the scoring test. If the user still does not meet internal quality control standards after completing the classroom course, their supervisor incorporates training into their performance evaluation plan.<sup>41</sup>

APD has instituted other mechanisms to ensure fidelity. In addition to mandatory initial and refresher training programs for presentence division staff and field probation officers (see *User Qualifications*), state presentence screeners are trained to perform quality control checks on the information gathered from the structured interview with the offender and entered into the automated system (such as by verifying criminal history information provided by the offender with existing records). Moreover, the computerized OST system automates the scoring process and contains built-in mechanisms to ensure that required questions are not skipped to minimize user error.

- **INSTRUMENT REVALIDATION.** In their 2008 independent evaluation, Lowenkamp and colleagues recommend that tests of the instruments' predictive

## Appendix: RNA Instrument Profile for the OST

---

validity should be conducted at least once every three years.<sup>42</sup>

### ENDNOTES

---

<sup>1</sup> See p. 4 in Arizona Adult Probation Services Division (2010, July). *Eight evidence-based principles for effective interventions & timeline of significant events in Arizona probation*. Phoenix, AZ: Author. Retrieved from

[http://www.azcourts.gov/LinkClick.aspx?fileticket=1CL48eRs\\_Nw%03d&tabid=2654](http://www.azcourts.gov/LinkClick.aspx?fileticket=1CL48eRs_Nw%03d&tabid=2654). Risk

principle discussed in Andrews, D. A., & Bonta, J. (2006). *The psychology of criminal conduct*, (4th ed.). Cincinnati: Anderson.

<sup>2</sup> See p. 474 in Ferguson, J. (2002). Putting the “what works” research into practice: An organizational perspective. *Criminal Justice and Behavior*, 29, 472-492. Retrieved from <http://cjb.sagepub.com/content/29/4/472>

<sup>3</sup> See pp. 474-475 in Ferguson (2002) at endnote 2.

<sup>4</sup> See endnote 1.

<sup>5</sup> See endnote 1.

<sup>6</sup> Lowenkamp, C. T., Latessa, E., & Bechtel, K. (2008). *A reliability and validation study of the Offender Screening Tool (OST) and Field Reassessment Offender Screening Tool (FROST) for Arizona*. Cincinnati: Center for Criminal Justice Research, University of Cincinnati.

<sup>7</sup> See Ferguson (2002) at endnote 2.

<sup>8</sup> See Lowenkamp et al. (2008) at endnote 6. See also Simourd, D. (2003). *Arizona Supreme Court: Administrative Office of the Courts, Adult Probation Services Division, Risk and needs assessment project*. Kingston, ON: Algonquin Correctional Evaluation Services.

<sup>9</sup> Baird, S.C. (1981). Probation and parole classification: The Wisconsin Model. *Corrections Today*, 43, 36-41.

<sup>10</sup> See Ferguson (2002) at endnote 2.

<sup>11</sup> Simourd, D. J., (2010, August). *Validation of the Offender Screening Tool (OST) for the Virginia local probation agencies*. Kingston,

---

ON: Algonquin Correctional Evaluation Services.

<sup>12</sup> See Ferguson (2002) at endnote 2.

<sup>13</sup> Arizona Adult Probation Services Division. (2009, July update). *FROST scoring guide*. Phoenix, AZ: Author.

<sup>14</sup> Arizona Adult Probation Services Division. (2009, July update). *MOST scoring guide*. Phoenix, AZ: Author.

<sup>15</sup> J. Ferguson and D. Simourd, personal communication, March 5, 2012.

<sup>16</sup> Andrews, D. A. & Bonta, J. (1995). *The Level of Service Inventory—Revised: User’s manual*. Toronto: Multi-Health System.

<sup>17</sup> See Ferguson (2002) at endnote 2.

<sup>18</sup> J. Ferguson and D. Simourd, personal communication, March 5, 2012..

<sup>19</sup> J. Ferguson and D. Simourd, personal communication, March 5, 2012.

<sup>20</sup> See Lowenkamp et al. (2008) at endnote 6.

<sup>21</sup> Arizona Adult Probation Services Division. (2009, July update). *OST scoring guide*. Phoenix, AZ: Author.

<sup>22</sup> J. Ferguson and D. Simourd, personal communication, March 5, 2012.

<sup>23</sup> See Lowenkamp et al. (2008) at endnote 6.

<sup>24</sup> See Ferguson (2002) at endnote 2.

<sup>25</sup> See Simourd (2003, July) at endnote 8.

<sup>26</sup> See Lowenkamp et al. (2008) at endnote 6.

<sup>27</sup> See Simourd (2010, August) at endnote 11.

<sup>28</sup> See Simourd (2003, July) at endnote 8.

<sup>29</sup> See Lowenkamp et al. (2008) at endnote 6.

<sup>30</sup> See Lowenkamp et al. (2008) at endnote 6.

<sup>31</sup> See Simourd (2010, August) at endnote 11.

<sup>32</sup> See Lowenkamp et al. (2008) at endnote 6.

<sup>33</sup> See Simourd (2003, July) at endnote 20.

<sup>34</sup> See Lowenkamp et al. (2008) at endnote 6.

<sup>35</sup> See Lowenkamp et al. (2008) at endnote 6.

<sup>36</sup> See Arizona Adult Probation Services Division (2009, July update) at endnote 13.

<sup>37</sup> See Lowenkamp et al. (2008) at endnote 6.

<sup>38</sup> J. Ferguson and D. Simourd, personal communication, March 5, 2012.

<sup>39</sup> J. Ferguson and D. Simourd, personal communication, March 5, 2012.

<sup>40</sup> J. Ferguson and D. Simourd, personal communication, March 5, 2012.

## Appendix: RNA Instrument Profile for the OST

---

---

<sup>41</sup> J. Ferguson and D. Simourd, personal communication, March 5, 2012.

<sup>42</sup> See Lowenkamp et al. (2008) at endnote 6.

# Ohio Risk Assessment System (ORAS): Community Supervision Tool (CST)

---

## ORAS GLOSSARY OF TERMS

---

<b>Risk</b>	Authors do not define “risk,” but explain the logic of Andrews and Bonta’s (1994) risk principle: The intensity of programmatic treatment should match the offender’s risk level so that “the most intensive programming should be allocated to moderate- and high-risk cases, while low-risk cases should be allocated little if any programming.” <sup>1</sup>
<b>Static risk</b>	Term not used in either report on the creation of the ORAS. However, authors use this term to describe risk factors that, because of the nature of the item(s), cannot be reduced over time. <sup>2</sup> Also referred to as “past criminal behavior.” <sup>3</sup>
<b>Dynamic risk</b>	Criminogenic or “crime-producing” needs, or “factors that, when changed, have been shown to result in a reduction in recidivism.” <sup>4</sup>
<b>Needs</b>	Authors do not use this term except to discuss criminogenic needs (see dynamic risk, above), but describe the needs principle as suggesting that “effective classification systems should identify dynamic risk factors directly related to recidivism so that they can be used to target programmatic needs.” <sup>5</sup>
<b>Responsivity</b>	Offender issues that “are not directly related to recidivism, but instead have the potential to restrict the efficacy of treatment. [They] are not used in the final calculation of risk, but instead are used as case planning factors that should be addressed to improve likelihood that programming will reduce recidivism.” <sup>6</sup>
<b>Protective factors</b>	Term not used in either report on the creation of the ORAS.
<b>Strengths</b>	Term not used in either report on the creation of the ORAS.
<b>Recidivism</b>	The ORAS Community Supervision Tool (CST) predicts the likelihood that community-based adult offenders will be arrested for a new crime, as measured in a 12-month follow-up period. <sup>7</sup>

---

## HISTORY & CURRENT USE.

---

**Creation.** In 2006, the Ohio Department of Rehabilitation & Correction (ODRC) hired researchers from the University of Cincinnati

(UC) Center for Criminal Justice Research to develop an integrated, automated assessment system of offender risk, needs, and barriers to treatment that could be used

## Appendix: RNA Instrument Profile for the ORAS-CST

---

to better inform decision-making statewide and ultimately reduce recidivism.<sup>8</sup>

**Current use.** The ODRC officially implemented the full ORAS statewide in Ohio as of March 2011, following the completion of construction, validation, and pilot testing studies on the system.<sup>9</sup> Although only recently adopted, a number of other states are using the ORAS, including Connecticut, Colorado, Montana, Nevada, New Hampshire, and Vermont, as well as a number of counties in Florida, Pennsylvania, and California.<sup>10</sup> A version of the complete ORAS was recently validated for statewide use in Indiana as the IRAS (i.e., Indiana Risk Assessment System) and in Texas as the TRAS (i.e., Texas Risk Assessment System).<sup>11</sup> Other studies are also currently planned or underway in Connecticut and Ventura County, California.<sup>12</sup>

### DEVELOPMENT.

---

**Instrument purpose.** The goal in creating the ORAS was to develop a unique, standardized system of offender assessment tools that could be used at various decision points in the criminal justice system to reduce recidivism, and that would facilitate communication and continuity in case management across criminal justice agencies.<sup>13</sup> The ORAS contains four full assessment tools (each designed for use at pretrial, at prison intake, with community supervision populations, or with reentry populations) and two brief screener tools (for use with prison and community supervision populations).<sup>14</sup> The authors have also recently developed a tool specifically for misdemeanants.<sup>15</sup> This profile focuses on the component of the ORAS developed

specifically for use with community-based populations of offenders (i.e., probation, parole, offenders in residential facilities or other community alternatives): the Community Supervision Tool, or CST. ORAS developers recommend administering the full CST, and not the short screening version, if using the results of the tool at the sentencing stage.<sup>16</sup>

**Approach to instrument development.** To create the CST, UC researchers adopted a prospective design.<sup>17</sup> This means that researchers identified current offenders (all adults charged with a criminal offense and referred to probation services during the period of data collection) for participation in the study, interviewed them to collect data on potential risk factors thought to predict recidivism, and observed these offenders over time (one year) to gather recidivism data. Researchers opted for a prospective study rather than a retrospective study which uses historical or archival data from past offenders to create the assessment tools because many potential offender risk factors considered for use in the CST or in other ORAS tools may not have been previously documented by criminal justice agencies. This approach allowed UC researchers to examine a comprehensive battery of over 200 potential risk factors for possible inclusion in the instrument(s).<sup>18</sup>

From this large pool of items, UC researchers eliminated those which failed to show a statistically significant relationship with recidivism. Researchers then conducted factor analyses and scale reliability tests to organize the content of the CST into seven domains or categories and to pare down the tool to the fewest items possible for optimal

## Appendix: RNA Instrument Profile for the ORAS-CST

---

predictive validity. In the item selection process, if a dynamic risk item performed as well or better than a comparable static risk item, UC researchers made a decision to prioritize the inclusion of the dynamic item because of the ability of dynamic items to measure and reflect changes over time. UC researchers indicated that generally, dynamic items were just as predictive as, if not better than, static risk items.<sup>19</sup>

### CONTENT.

---

**Structure.** The CST generates a single overall score from a set of seven subscales. This overall score represents the offender's risk of recidivism. Scores for each of the seven separate subscales of the CST are used to identify and prioritize the offender's needs for case planning and service provision (see *Items and domains* section below for a list of the needs domains addressed by the CST).

A separate section lists responsivity factors as other potential areas of concern that may inform case planning decisions. These factors are not criminogenic and are not incorporated into the computation of risk.<sup>20</sup>

**Items and domains.** The ORAS CST consists of 35 items in 7 subscales: criminal history (6 items); education, employment, and finances (6 items); family and social support (5 items); neighborhood problems (2 items); substance abuse (5 items); antisocial associations (4 items); and antisocial attitudes and behavioral problems (7 items).

The CST also documents the following treatment barriers to inform case planning: low intelligence, physical handicap, reading and writing limitations, mental health issues,

offender motivation to change/participate in treatment, transportation, child care, language, ethnicity, cultural barriers, history of abuse/neglect, and interpersonal anxiety.<sup>21</sup>

**Reporting and cutoffs.** The ORAS CST groups offenders into four levels of risk (low, moderate, high, very high) based on their overall score. The cutoff scores differ by gender.

The ORAS CST also groups offenders, on each subscore, into three priority levels (low, moderate, high) to inform decisions about which offender needs should be prioritized in case planning and service provision. Offenders categorized as "high" in a particular domain are more likely to reoffend. The cut points vary by domain, but not by gender.

All cutoff scores are identified in the ORAS manual.<sup>22</sup>

### INSTRUMENT RELIABILITY AND VALIDITY.

---

**Populations studied.** In addition to the statewide Ohio creation and validation samples of probation-eligible male and female adult offenders, Indiana has also completed a statewide validation study of the tool (report forthcoming).<sup>23</sup>

**Predictive validity.** ORAS developers reported a correlation of  $r = .36$  between ORAS CST risk level and recidivism in the Ohio study. Moreover, case management priority levels for each of the 7 subscale domains also correlated individually with recidivism (criminal history,  $r = .20$ ; education and finances,  $r = .22$ ; social support,  $r = .12$ ; neighborhood problems,  $r =$

## Appendix: RNA Instrument Profile for the ORAS-CST

---

.20; substance abuse,  $r = .14$ ; antisocial associates,  $r = .32$ ; and antisocial attitudes,  $r = .24$ ), providing further evidence that these domains identify criminogenic needs.<sup>24</sup>

**Reliability.** No data available at the time of this report but see “Independent validation” section below.

**Potential for bias.** There is little evidence currently available on the issue of bias with the ORAS CST.

- **GENDER.** ORAS developers reported correlations between the ORAS CST risk level and recidivism in the Ohio study of  $r = .37$  for males and  $r = .30$  for females.<sup>25</sup> In general, female offenders tend to produce lower scores on the ORAS than males. Instrument developers established different risk level cutoff scores by gender to reflect this.<sup>26</sup>
- **RACE.** No data currently available.

**Independent validation.** As of this publication, no independent validation studies of the ORAS have been published. However, Texas reportedly has recently completed the first independent interrater reliability study and an independent predictive validity study using a random statewide sample.<sup>27</sup>

### PRACTICAL CONSIDERATIONS.

---

**Vendor and instrument cost.** The ORAS tools are non-proprietary. For more information, contact Ms. Jennifer Luxat UC ([luxjl@ucmail.uc.edu](mailto:luxjl@ucmail.uc.edu)).

**Menu of other services.** UC offers a wide array of services, training, and technical assistance to support ORAS implementation.

- **IT SERVICES.** Customized software is available for purchase. Depending on the level of customization and other options selected, the price of an automated module system currently ranges from \$15K – 100K.<sup>28</sup> As of this report, customization options include:<sup>29</sup>
  - A base module system that is hosted on the UC server
  - A customized module system with client branding that is hosted on the UC server
  - A customized module system with client branding that is hosted on the UC server, but that allows data sharing from the UC server to the client through specialized web services or file transfers
  - A customized module system with client branding that is hosted on the client server
  - A customized module system with client branding that is hosted on the client server and that is either (a) integrated into the existing case management system or (b) is a stand-alone system that allows information sharing with other existing systems on the client server.
- **VALIDATION SERVICES.** With the ORAS, clients retain the rights to their own data. Clients may choose to (a) conduct the validation analysis in-house, (b) send the data out to an external reviewer for validation, or (c) hire UC to perform the validation analysis.<sup>30</sup>
- **USER TRAINING.** As of this report, UC provides a 2-day basic ORAS training for \$7000, including trainer travel expenses.<sup>31</sup>

## Appendix: RNA Instrument Profile for the ORAS-CST

---

Also offered is a “train the trainer” course for those agencies that are interested in developing the internal capacity to sustain the use of the ORAS.<sup>32</sup> For more information about other training services offered by UC, contact Mr. John Schwartz at [John.Schwartz@uc.edu](mailto:John.Schwartz@uc.edu) or visit their website: <http://www.uc.edu/corrections/services/trainings.html>.

**User qualifications.** The basic user training is mandatory. This includes an overview of the ORAS tools, training on the techniques for administering and scoring individual assessments, and training on how to use the ORAS in case management.<sup>33</sup>

**Administration time.** The ORAS CST takes approximately 50 minutes to administer.<sup>34</sup>

**Modes of administration.** Information collected to complete the ORAS CST is obtained through a structured interview with the offender and an offender self-report form. Assessors are encouraged to corroborate information whenever possible with official records and collateral sources.<sup>35</sup>

**Quality assurance.** When adopting any offender assessment tool, jurisdictions must be prepared to ensure appropriate implementation and proper maintenance over time. Quality assurance recommendations and guidelines for the ORAS CST follow.<sup>36</sup>

- **OVERRIDE POLICY.** Generally, overrides may occur if (a) the user determines that the risk assessment does not reflect the actual risk of the offender and wishes to change the assessed risk level in the individual case, or (b) if, given the

assessed risk of the offender, the user must override for policy reasons (e.g., a mandate to place a particular type of offender in maximum supervision regardless of assessed risk level). ORAS developers recommend an override rate of 2-3% or less; however, overrides should not occur in more than 10% of the total population of cases and, for an individual assessor, in more than 10% of his or her caseload. If judges receive ORAS CST results, they should be notified of any override.

- **FIDELITY.** The ORAS CST interview guide is structured to increase reliability between assessors. Moreover, the automated system provided by UC includes program and data sharing features that can help minimize assessor error. However, as with any offender assessment tool, routine fidelity studies of the ORAS CST are recommended. For this purpose, the automated system includes a feature which allows the client to draw a random sample of cases (5-10%) for internal review. Clients can seek certification training from UC to learn how to conduct these studies internally.
- **INSTRUMENT REVALIDATION.** UC researchers recommend that clients revalidate the ORAS tool(s) approximately every five years.

### ENDNOTES

---

<sup>1</sup> See p. 16 in Latessa, E. J., Lemke, R., Makarios, M., Smith, P., & Lowenkamp, C. T. (2010). The creation and validation of the Ohio Risk Assessment System (ORAS)\*. *Federal Probation*, 74, 16-22. Retrieved from

## Appendix: RNA Instrument Profile for the ORAS-CST

---

[http://www.uscourts.gov/uscourts/FederalCourts/PPS/Fedprob/2010-06/02\\_creation\\_validation\\_of\\_oras.html](http://www.uscourts.gov/uscourts/FederalCourts/PPS/Fedprob/2010-06/02_creation_validation_of_oras.html)

<sup>2</sup> Brian Lovins, personal interview, February 16, 2012.

<sup>3</sup> See slide 22 in Lovins, B. (2010, September 22). *The development and implementation of the Indiana Risk Assessment System* [PowerPoint slides]. Center for Criminal Justice Research, University of Cincinnati.

<sup>4</sup> See p. 16 in Latessa et al. (2010) at endnote 1 and slide 22 in Lovins (2010) at endnote 3.

<sup>5</sup> See p. 8 in Latessa, E., Smith, P., Lemke, R., Makarios, M., & Lowenkamp, C. (2009). *Creation and validation of the Ohio Risk Assessment System: Final report*. Cincinnati, OH: University of Cincinnati Center for Criminal Justice Research. Retrieved from [http://www.ocjs.ohio.gov/ORAS\\_FinalReport.pdf](http://www.ocjs.ohio.gov/ORAS_FinalReport.pdf)

<sup>6</sup> See p. 18 in Latessa et al. (2010) at endnote 1.

<sup>7</sup> See Latessa et al. (2010) at endnote 1.

<sup>8</sup> See Latessa, et al. (2009) at endnote 5.

<sup>9</sup> Ohio Department of Rehabilitation and Correction (2010, June 15), *Ohio Risk Assessment System*. Retrieved from <http://www.drc.ohio.gov/web/ORAS.htm>

<sup>10</sup> Counties include Ventura, Riverside and Orange in California, Alachua, Orange, and Seminole in Florida, and Dauphin, Berks, and York in Pennsylvania.

<sup>11</sup> E. Latessa, personal communication, June 11, 2014. An IRAS validation report is forthcoming. Brian Lovins, personal interview, February 16, 2012. See Goodman, M. & Thompson, L. (2011, April 13). Indiana's new risk assessment tools: What you should know. *Indiana Court Times*, 20.2, 5-7. Retrieved from

<http://issuu.com/incourts/docs/news20-2c?mode=window&backgroundColor=%02322222>. See also Lovins (2010) at endnote 3.

<sup>12</sup> B. Lovins, personal communication, February 16, 2012.

<sup>13</sup> See Latessa, et al. (2010) at endnote 1.

<sup>14</sup> See Latessa et al. (2009) at endnote 5. ORAS developers strongly recommend that these screener tools be carefully validated at the local level prior to use in other jurisdictions.

<sup>15</sup> E. Latessa, personal communication, June 11, 2014. See Latessa, E., Lovins, B., & Lux, J. (2014). *The Ohio Risk Assessment System Misdemeanor Assessment Tool (ORAS-MAT) and Misdemeanor Screening Tool (ORAS-MST)*. Cincinnati: University of Cincinnati Center for Criminal Justice Research.

<sup>16</sup> B. Lovins, personal communication, February 16, 2012.

<sup>17</sup> See Latessa et al. (2009) at endnote 5.

<sup>18</sup> See Latessa et al. (2009) at endnote 5; also B. Lovins, personal communication, February 16, 2012.

<sup>19</sup> *Id.*

<sup>20</sup> Center for Criminal Justice Research, University of Cincinnati (n.d.). *Ohio Risk Assessment System* (user manual). Cincinnati, OH: Authors.

<sup>21</sup> *Id.*

<sup>22</sup> *Id.*

<sup>23</sup> B. Lovins, personal communication, February 16, 2012.

<sup>24</sup> See Latessa et al. (2009) at endnote 5.

<sup>25</sup> See Latessa et al. (2010) at endnote 1. To assess the predictive validity of the ORAS CST, Receiver Operating Characteristics (ROC) analyses were also performed, producing Area Under the Curve (AUC) values of .71 for males and .69 for females.

<sup>26</sup> See Latessa et al. (2010) at endnote 1.

<sup>27</sup> E. Latessa, personal communication, June 11, 2014.

<sup>28</sup> E. Latessa, personal communication, June 11, 2014.

<sup>29</sup> University of Cincinnati (n.d.). *Ohio Risk Assessment System web-based modules*. Cincinnati, OH: Authors.

<sup>30</sup> B. Lovins, personal communication, February 16, 2012.

<sup>31</sup> B. Lovins, personal communication, February 16, 2012..

## Appendix: RNA Instrument Profile for the ORAS-CST

---

---

<sup>32</sup> B. Lovins, personal communication,  
December 7, 2012.

<sup>33</sup> B. Lovins, personal communication,  
December 7, 2012.

<sup>34</sup> B. Lovins, personal communication,  
December 7, 2012.

<sup>35</sup> B. Lovins, personal communication,  
December 7, 2012.

<sup>36</sup> B. Lovins, personal communication,  
December 7, 2012.

# The Static Risk and Offender Needs Guide (STRONG)

---

## STRONG GLOSSARY OF TERMS

---

<b>Risk</b>	Term is not explicitly defined in published sources on the STRONG. However, sources do refer to Andrews and Bonta's (1994) risk principle.
<b>Static risk</b>	"Risk factors that cannot decrease, such as criminal history, are static. Once a criminal record is obtained, it will always be a part of an offender's history" (Barnoski & Drake, 2007, p. 2; citing Andrews & Bonta, 1998).
<b>Dynamic risk</b>	"Dynamic risk factors, such as drug dependency, can decrease through treatment or intervention" (Barnoski & Drake, 2007, p. 2; citing Andrews & Bonta, 1998).
<b>Needs</b>	Term is not explicitly defined, but sources refer to these as dynamic, criminogenic factors that may be addressed in re-entry and supervision planning.
<b>Responsivity</b>	Term not used by STRONG developers or vendor.
<b>Protective factors</b>	Term is not explicitly defined, but sources refer to these as factors that, when present or when increased, can reduce recidivism.
<b>Strengths</b>	Term not used by STRONG developers or vendor.
<b>Recidivism</b>	The state of Washington defines recidivism as "a subsequent conviction in a Washington State Superior Court for a felony offense committed within three years of placement in the community. In addition, one year is allowed for the offense to be adjudicated in court" (Barnoski & Drake, 2007, p. 2). The static risk assessment component of the STRONG system predicts felony recidivism and distinguishes between high drug, property, and violent felony risk.

---

## HISTORY & CURRENT USE.

---

**Creation.** In 1999, the Washington State Legislature passed the Offender Accountability Act (effective July 2000), which called for improved efforts to "reduce the risk of reoffending by offenders in the community" (RCW 9.94A.010). The Washington State Institute for Public Policy

(WSIPP) was charged with evaluating the impact of these legislative changes on recidivism. In a 2003 report, WSIPP recommended improvements to the predictive accuracy of the Washington State Department of Corrections' (DOC) previous assessment tool (the LSI-R) by including more static risk items in the assessment.<sup>1</sup> The DOC requested that WSIPP create a new static risk assessment instrument comprised

## Appendix: RNA Instrument Profile for the STRONG

---

entirely of criminal history and demographic items and a new needs assessment instrument of offender deficits and protective factors for statewide use.<sup>2</sup>

WSIPP researchers developed the Static Risk Assessment in 2006 and created the Offender Needs Assessment to complete the STRONG system.<sup>3</sup> Assessments.com collaborated with a DOC team to build a software application for the STRONG and integrated it with the existing state case management system.<sup>4</sup> The STRONG was fully implemented by the Washington State DOC in August 2008.<sup>5</sup>

**Current use.** In addition to Washington State where the STRONG was developed and has been in use since 2008, the system has also reportedly been used by multiple jurisdictions in California (over 30 counties), Florida, and Texas.<sup>6</sup>

### DEVELOPMENT.

---

**Instrument purpose.** The Static Risk Assessment was designed for statewide use to assess offenders' recidivism risk, and the Offender Needs Assessment was developed to identify dynamic offender needs and protective factors that can be addressed in reentry and supervision planning.<sup>7</sup>

The Washington State DOC chose to develop static risk and offender needs assessments over the tool they previously used to assess offender risk and needs. This decision was based on the results of a WSIPP validation study on the previously used instrument. The DOC listed a number of reasons for this decision, including the increased accuracy of risk prediction in the

state with the Static Risk Assessment; greater specificity in prediction by classifying high risk offenders according to the most serious type of crime predicted (drug, property, violent); increased objectivity of a tool that is based on verifiable demographic and criminal history data rather than questions from structured interviews; decreased costs associated with the administration of the tool; and more accurate documentation of criminal history information for use in other DOC applications.<sup>8</sup>

### **Approach to instrument development.**

WSIPP researchers adopted a retrospective design in creating the Static Risk Assessment.<sup>9</sup> This means that researchers identified a "construction sample" of offenders (in this case, all 308,423 offenders released from incarceration or placed on community supervision in Washington State from 1986 to March of 2000) and used archival offender and felony reconviction data to determine which demographic and criminal history factors were most strongly associated with recidivism. Researchers applied multivariate regression techniques to identify variables that most strongly predicted recidivism for inclusion in the Static Risk Assessment tool and to develop a weighted algorithm for the calculation of risk scores. WSIPP researchers then validated this Static Risk Assessment on a sample of 51,648 Washington State felony offenders who were released from incarceration or placed on community supervision from 2001 through September 2002.

The Offender Needs Assessment was developed through a collaborative effort between WSIPP and a focus group of state correctional officers. The tool contains

## Appendix: RNA Instrument Profile for the STRONG

---

dynamic items that have a demonstrated relationship with recidivism in the broader scientific literature and also some non-criminogenic items identified by correctional officers as important for case management.<sup>10</sup> No published documentation is yet available on the development or validation of the Offender Needs Assessment tool.

### CONTENT.

---

**Structure.** The STRONG consists of two separate assessment instruments: the Static Risk Assessment and the Offender Needs Assessment. The Static Risk Assessment is designed to assess offender risk for reoffense and classify each offender to a single risk category for case management purposes. It is used to determine the amount of supervision the offender receives and the prioritization for services. Recently, Washington State conducted a study to assess the feasibility of implementing the Static Risk Assessment as a standard assessment in seven state court pretrial programs to inform pretrial release and alternative sentencing decisions. As of this report, researchers are developing a modified version of the Static Risk Assessment for use by the courts at pretrial statewide.<sup>11</sup>

The separate Offender Needs Assessment is designed to identify offender deficits and protective factors for use in guiding decisions about the type of service programming that would be most appropriate. This assessment includes dynamic criminogenic factors as well as static and non-criminogenic items identified by correctional officers as relevant to professional judgment in case planning.<sup>12</sup>

**Items and domains.** The Static Risk Assessment component of the STRONG collects information on 26 items in 6 general categories: demographic information (2 items), juvenile felony convictions and commitments (4 items), DOC commitments (1 item), felony conviction types (9 items), misdemeanor conviction types (9 items), and adult sentence violations (1 item).<sup>13</sup>

The Offender Needs Assessment component of the STRONG system in Washington State collects information on 55 items across 10 gender-neutral domains related to criminal behavior: education (4 items), community employment (10 items), friends (2 items), residential (3 items), family (8 items), alcohol and drug use (6 items), mental health (6 items), aggression (4 items), attitudes and behaviors (7 items), and coping skills (5 items). These domains assess offender needs and protective factors supported by “best practices” in the broader social learning research literature as related to criminal behavior. These factors include the presence of antisocial associates and absence of prosocial others (community employment, friends, family domains); attitudes, values, and beliefs supportive of criminal behaviors (aggression, attitudes/behaviors, coping skills domains); personality traits (alcohol/drug use, mental health, aggression domains); personal achievement (education, community employment, residential domains); and family dynamics (family domain).<sup>14</sup>

**Reporting and cutoffs.** The Static Risk Assessment groups offenders into five levels of risk (low, moderate, high drug, high property, high violent). Three separate weighted algorithms are used to compute

## Appendix: RNA Instrument Profile for the STRONG

---

general felony risk, property felony risk, and violent felony risk; these risk score calculations are used to determine the offender's classification of risk.<sup>15</sup> The Washington State DOC subsequently revised the five offender classification levels down to four groupings (low, moderate, high non-violent [property, drug], high violent).<sup>16</sup>

The Offender Needs Assessment identifies whether each of the 10 domains is considered a low, moderate, or high need and/or a low, moderate, or high protective factor. The greater the need, the more of a priority the domain is in case planning.<sup>17</sup>

### INSTRUMENT RELIABILITY AND VALIDITY.

---

**Populations studied.** The Static Risk Assessment construction and validation samples included adult community supervision and prison cohort groups in Washington State. Men and women were represented within these samples, as were various racial groups (European, African, Native, Asian, and Hispanic Americans) and types of offenses (drug, property, sex, violent non-sex offenses).<sup>18</sup> No data is yet available on the Offender Needs Assessment.

**Predictive validity.** To assess the predictive validity of the Static Risk Assessment, Receiver Operating Characteristic (ROC) analyses were performed on construction and validation samples of Washington state felony offenders on community supervision or in prison. These studies produced Area Under the Curve (AUC) values of .756 and .742 for these two samples.<sup>19</sup> No data is yet available on the validity of the Offender Needs Assessment.

**Reliability.** No data yet available.

**Potential for bias.** In the initial validation study of the Static Risk Assessment, WSIPP researchers examined the efficacy of the tool by gender and race.<sup>20</sup>

- **GENDER** The Static Risk Assessment discriminates equally well by gender (for felony offenses generally, among males, AUC = .743; among females, AUC = .720). However, Barnoski and Drake explain that the tool tends to underestimate property recidivism and overestimate violent recidivism for females compared to males.
- **RACE.** The tool also discriminates well by racial group (for felony offenses generally, among European Americans, AUC = .736; among African Americans, AUC = .723; among Native Americans, AUC = .716; among Asian Americans, AUC = .748; and among Hispanic Americans, AUC = .742). However, Barnoski and Drake explain that the tool seems to perform less well for Asian Americans in discriminating between high drug and high property recidivism.

In addition, the tool predicts violent recidivism for sex offenders but not sexual reoffending.<sup>21</sup> No data is yet available on the Offender Needs Assessment.

**Independent validation.** As of this publication, no validation studies of the STRONG have yet been published by a research organization independent from WSIPP or the Washington State DOC.

## Appendix: RNA Instrument Profile for the STRONG

---

### PRACTICAL CONSIDERATIONS.

---

**Vendor and instrument cost.** The STRONG instruments are non-proprietary. However, the STRONG software application is proprietary. Software programs for the STRONG and custom integration services are currently offered by two companies: Noble Software Group and Assessments.com. For more information on Noble Software Group, contact [info@noblesg.com](mailto:info@noblesg.com) or call (979) 248-6568. To contact an Assessments.com representative, email [info@assessments.com](mailto:info@assessments.com) or call 877-277-3778.

**Menu of other services.** Both Assessments.com and Noble Software Group offer an array of training, technical assistance, and other services to support the implementation of the STRONG.

- **IT SERVICES.** Both companies offer two general approaches to STRONG software implementation:
  - A hosted solution on remote servers for a recurring fee; may require a set-up fee
  - An enterprise solution on the agency's own servers; licensed software will run in-house, with or without customized integration.Both companies offer custom report generation and an automated case plan software product to help users build individual case plans from information on offenders' needs. Pricing is established based on the number of user licenses, not the number of assessments or reassessments.<sup>22</sup>
- **VALIDATION SERVICES.** Both companies recommend local validation of the

STRONG prior to implementation and will employ consultants to assist in this process if requested by the client. Pricing is determined based on the number of consulting hours required to conduct the validation study.<sup>23</sup>

- **USER TRAINING.** Noble Software Group and Assessments.com offer a two-day training on the STRONG, which is required before staff may use the tool. It is also strongly recommended that staff attend a two-day training on motivational interviewing before using the Offender Needs Assessment, and that users attend a booster training to enhance their skill set after they have used the STRONG for a few months.<sup>24</sup> In Washington State, probation officers in the field are trained in motivational interviewing techniques prior to conducting an Offender Needs Assessment interview.<sup>25</sup> Other training, including Train-the-Trainer programs, are also offered by both companies. As of this report, trainings typically cost approximately \$2,500.00 per day from Assessments.com.<sup>26</sup> Visit [https://www.assessments.com/content/training\\_curricula.asp](https://www.assessments.com/content/training_curricula.asp) or contact Assessments.com for more information. Trainings typically cost \$2,200 per day from Noble Software Group. Visit <http://www.noblesg.com> for more information on Noble's training programs.

**User qualifications.** The two-day STRONG training is mandatory for all users.

**Administration time.** The Static Risk Assessment component of the STRONG can

---

## Appendix: RNA Instrument Profile for the STRONG

---

take up to 15-30 minutes per offender, depending on the complexity of the offender's criminal history.<sup>27</sup> The Offender Needs Assessment takes approximately 1 hour.<sup>28</sup>

**Modes of administration.** The Static Risk Assessment is based on criminal history and demographic data extracted from case files. In Washington State, the Static Risk Assessment is conducted by a specialized, centralized unit of 13 officers<sup>29</sup> with access to out-of-state criminal history information from the Washington State Justice Information System and the National Crime Information Center. Some jurisdictions, however, opt to auto-populate the Static Risk Assessment using information from their existing management information systems.<sup>30</sup>

The Offender Needs Assessment is completed with information gathered by the probation officer from a file review, a structured interview with the offender, and collateral contacts.<sup>31</sup> Scores are automatically computed in the software application and reports are automatically generated.

**Quality assurance.** When adopting any offender assessment tool, jurisdictions must be prepared to ensure appropriate implementation and proper maintenance over time. Quality assurance recommendations and guidelines for the STRONG follow.

- **VERRIDE POLICY.** The need for an override is determined by the probation officer on a case-by-case basis and as guided by local policy. The Washington State DOC has reportedly observed a 5-10% exception rate with the tool.<sup>32</sup>
- **FIDELITY.** Assessments.com does not provide quality assurance standards for the STRONG per se. Rather, they recommend a comprehensive approach in which local implementation teams are assembled, with input from research consultants, to facilitate local decision-making about necessary business rules and continuous quality improvement needs.<sup>33</sup> Noble Software Group provides additional inter-rater reliability software products as part of a quality assurance process to ensure long-term fidelity to the instruments.  
  
In Washington State, the DOC employs trained subject matter experts who conduct routine quality assurance testing. These efforts involve observations of offender interviews and reviews of completed assessments. Additional peer support meetings and training are provided for offices struggling with quality control issues.<sup>34</sup> WSIPP developers recommend good initial training and some form of regular case review round table meetings within each unit to address quality assurance issues and to encourage ongoing dialogue about how STRONG information may be appropriately used in case management/planning.<sup>35</sup>
- **INSTRUMENT REVALIDATION.** The instrument's developer, Robert Barnoski, has indicated that the frequency of revalidation depends in part on how the instrument is used.<sup>36</sup> In Washington, the predictive accuracy of the tool is monitored annually to determine whether or not the recidivism rates

## Appendix: RNA Instrument Profile for the STRONG

---

within each risk classification level remain fairly constant. If the rates remain constant, revalidation may not be necessary. However, if evidence arises that the tool is no longer working appropriately or if significant policy changes affect the ability to use the tool as originally intended, Dr. Barnoski recommends conducting a revalidation study.

### ENDNOTES

---

<sup>1</sup> Barnoski, R., & Aos, S. (2003). *Washington's Offender Accountability Act: An analysis of the Department of Corrections' risk assessment*. Olympia, WA: Washington State Institute for Public Policy.

<sup>2</sup> Barnoski, R., & Drake, E., (2007). *Washington's Offender Accountability Act: Department of Corrections' Static Risk Instrument*. Olympia, WA: Washington State Institute for Public Policy.

<sup>3</sup> *Id.*

<sup>4</sup> Assessments.com (2011). *Washington State Department of Corrections implements the STRONG*. Bountiful, UT: Authors.

<sup>5</sup> M. Miller, personal communication, August 25, 2010.

<sup>6</sup> R. Barnoski, personal communication, April 24, 2012; also C. Lake, personal communication, August 23, 2012.

<sup>7</sup> Washington State Department of Corrections (n.d.). *Offender Needs Assessment help text*. Olympia: Washington.

<sup>8</sup> M. Miller, personal communication, August 25, 2010. See also Drake, E., & Barnoski, R. (2009). *New risk instrument for offenders improves classification decisions*. Olympia: Washington State Institute for Public Policy.

<sup>9</sup> See Barnoski & Drake (2007) at endnote 2.

<sup>10</sup> R. Barnoski, personal communication, April 24, 2012.

<sup>11</sup> Washington State Administrative Office of the Courts Information Services Division (2011). *Adult Static Risk Assessment*

feasibility study. Olympia, WA: Washington State Administrative Office of the Courts.

<sup>12</sup> R. Barnoski, personal communication, April 24, 2012.

<sup>13</sup> See Barnoski & Drake (2007) at endnote 2.

<sup>14</sup> Washington State Department of Corrections (2007). *Offender Needs Assessment tool* (factoids and weights). Olympia, WA: Authors. See also R. Barnoski, personal communication, April 24, 2012.

<sup>15</sup> See Barnoski & Drake (2007) at endnote 2 for detailed information about the classification rules and score cutoffs used to determine risk level.

<sup>16</sup> M. Miller, personal communication, August 25, 2010.

<sup>17</sup> Washington State Department of Corrections (2007) at endnote 14. See also Assessments.com (2011) at endnote 4.

<sup>18</sup> See Barnoski & Drake (2007) at endnote 2.

<sup>19</sup> *Id.*

<sup>20</sup> *Id.*

<sup>21</sup> *Id.*

<sup>22</sup> C. Lake, personal communication, August 2, 2012.

<sup>23</sup> C. Lake, personal communication, August 2, 2012.

<sup>24</sup> C. Lake, personal communication, August 2, 2012.

<sup>25</sup> M. Miller, personal communication, August 25, 2010.

<sup>26</sup> C. Lake, personal communication, August 23, 2012.

<sup>27</sup> Washington State Administrative Office of the Courts Information Services Division (2011) at endnote 11.

<sup>28</sup> R. Barnoski, personal communication, April 24, 2012.

<sup>29</sup> M. Miller, personal communication, August 25, 2010.

<sup>30</sup> R. Barnoski, personal communication, April 24, 2012.

<sup>31</sup> Washington State Department of Corrections (n.d.) at endnote 7.

<sup>32</sup> M. Miller, personal communication, August 25, 2010.

## Appendix: RNA Instrument Profile for the STRONG

---

---

<sup>33</sup> C. Lake, personal communication, August 2, 2012 and August 23, 2012.

<sup>34</sup> Assessments.com (2011) at endnote 4.

<sup>35</sup> R. Barnoski, personal communication, April 24, 2012.

<sup>36</sup> *Id.*