



## ARTICLE

# What Even Is a Criminal Attitude?—And Other Problems with Attitude and Associational Factors in Criminal Risk Assessment

Beth Karp\*

**Abstract.** Several widely used criminal risk assessment instruments factor a defendant's abstract beliefs, peer associations, and family relationships into their risk scores. The inclusion of those factors is empirically unsound and raises profound ethical and constitutional questions. This Article is the first instance of legal scholarship on criminal risk assessment to (a) conduct an in-depth review of risk assessment questionnaires, scoresheets, and reports, and (b) analyze the First and Fourteenth Amendment implications of attitude and associational factors. Additionally, this Article challenges existing scholarship by critiquing widely accepted but dubious empirical justifications for the inclusion of attitude and associational items. The items are only weakly correlated with recidivism, have not been shown to be causal, and have in fact been shown to decrease the predictive validity of risk assessment instruments. Quantification of attitudes and associations should cease unless and until it is done in a way that is empirically sound, more useful than narrative reports, and consistent with the First and Fourteenth Amendments.

---

\* Attorney-Advisor, U.S. Board of Veterans' Appeals; J.D., University of Chicago; M.P.P., University of Chicago. Thanks to Aziz Huq for guidance throughout the development and writing process; to Barbara Barreno-Paschal, David Karp, and Swethaa Ballakrishnen for their valuable comments; and to Ninet Sinaii for advice on empirical analysis. My deepest gratitude, as well, to the editors of the *Stanford Law Review* for their indispensable editorial work.

## Table of Contents

|   |      |
|---|------|
| Introduction .....  | 1434 |
| A. Background .....   | 1434 |
| B. Article Overview .....   | 1437 |
| I. Overview of Select Risk Assessment Instruments .....   | 1437 |
| A. Correctional Offender Management Profiling for Alternative Sanctions<br>(COMPAS).....          | 1439 |
| B. Level of Service Inventory–Revised (LSI-R).....  | 1444 |
| C. Ohio Risk Assessment System (ORAS) and Its Derivatives .....                                   | 1451 |
| D. Offender Screening Tool (OST) and Modified Offender Screening Tool<br>(MOST).....              | 1454 |
| 1. OST and MOST content—known and inferred.....   | 1454 |
| 2. OST and MOST design problems.....  | 1465 |
| II. How Risk-Needs Quantification Works and Doesn’t Work.....                                     | 1468 |
| A. How Risk is Measured.....  | 1468 |
| 1. Typical risk assessment factors .....  | 1468 |
| 2. Commonly used statistics: AUC and <i>r</i> .....   | 1470 |
| a. AUC.....   | 1470 |
| b. Correlation coefficients ( <i>r</i> ).....   | 1471 |
| B. No One Has a Good Handle on How to Assess Criminogenic Needs.....                              | 1476 |
| 1. AUC and <i>r</i> tell an incomplete story .....  | 1476 |
| 2. The Gendreau meta-analysis is not an oracle.....   | 1479 |
| 3. Dynamic factors are only weakly correlated with crime.....                                     | 1484 |
| 4. Call for item-level transparency.....  | 1487 |
| C. More Factors, More Problems.....   | 1489 |
| 1. Attitude and associational factors do not improve predictive<br>accuracy.....                  | 1489 |
| 2. “Need” factors should be explained, not quantified .....                                       | 1492 |
| III. Speech and Expressive Association.....   | 1494 |
| A. Abstract Beliefs are Inadmissible at Sentencing Under the First<br>Amendment .....             | 1494 |
| 1. <i>Dawson</i> requires a nexus.....  | 1494 |
| 2. Factors in general risk assessment instruments lack a nexus to the<br>defendant’s offense..... | 1498 |
| 3. “It’s Just One Factor” does not cure First Amendment harms .....                               | 1498 |
| B. Normative Issues in Attitude Assessment.....   | 1500 |
| 1. What even is a criminal attitude?.....   | 1500 |
| 2. Attitudes as racial proxies.....   | 1503 |
| 3. Attitude measurement problems.....   | 1506 |
| IV. Peer and Family Relationships.....  | 1508 |
| A. Risk Assessment Associational Items .....  | 1509 |

*What Even Is a Criminal Attitude?*  
75 STAN. L. REV. 1431 (2023)

|  |      |
|--|------|
| 1. Criminal peers.....   | 1509 |
| 2. Family criminal history.....  | 1510 |
| B. Constitutional Protections for Peer and Family Relationships.....                             | 1512 |
| 1. The substantive due process dead end.....   | 1513 |
| a. The <i>Jaycees</i> quagmire.....  | 1513 |
| b. Peer association items do not violate substantive due process .....                           | 1515 |
| 2. The First Amendment potential pathway .....   | 1516 |
| 3. The Equal Protection proscription of kin punishment.....                                      | 1519 |
| C. Echoes of Eugenics in Criminal Risk Assessment.....   | 1523 |
| D. Garbage-In-Garbage-Out-Garbage-In-Garbage-Out (Peer Association<br>Measurement Problems)..... | 1524 |
| V. Beware the Shiny Legal Penny .....  | 1527 |
| Conclusion.....  | 1528 |

## Introduction

### A. Background

Risk assessment instruments have become pervasive in American criminal justice, heralded as a means to reduce mass incarceration, protect the public, conserve resources, and properly individualize sentencing.<sup>1</sup> The instruments are intended to supplement human discretion in criminal justice decisionmaking<sup>2</sup> by taking a set of quantitative inputs, like number of prior convictions or level of educational attainment,<sup>3</sup> and attempting to predict the risk of outcomes like recidivism or failure to appear.<sup>4</sup> Their most well-known and controversial usage is in pretrial release determinations, but they are employed at every phase of criminal justice proceedings, including sentencing, prison assignment, and parole.<sup>5</sup> Nearly every state uses a predictive instrument in at least one phase of criminal proceedings, as does the federal government.<sup>6</sup>

- 
1. See MODEL PENAL CODE: SENT'G § 6B.09 cmt. d (AM. L. INST., Proposed Official Draft 2017) (“If used as a tool to encourage sentencing judges to divert low-risk offenders from prisons to community sanctions, risk assessments conserve scarce prison resources for the most dangerous offenders, reduce the overall costs of the corrections system, and avoid the human costs of unneeded confinement to offenders, offenders’ families, and communities.”); SARAH L. DESMARAIS & EVAN M. LOWDER, SAFETY & JUST. CHALLENGE, PRETRIAL RISK ASSESSMENT TOOLS: A PRIMER FOR JUDGES, PROSECUTORS, AND DEFENSE ATTORNEYS 3 (2019), <https://perma.cc/EB75-G8WF> (“The overarching reform vision is to shift from the ‘resource-based’ system of money bail to a ‘risk-based’ system, in which pretrial interventions are tied to risk rather than wealth.”).
  2. PAMELA M. CASEY, JENNIFER K. ELEK, ROGER K. WARREN, FRED CHEESMAN, MATT KLEIMAN & BRIAN OSTROM, NAT’L CTR. FOR STATE CTS., OFFENDER RISK & NEEDS ASSESSMENT INSTRUMENTS: A PRIMER FOR COURTS 2 (2014) (“Practitioners use risk assessment information to inform decisions at various points in the criminal justice system.”).
  3. See Level of Service Inventory–Revised (IDOC) (n.d.), <https://perma.cc/R4TJ-VUEQ>. This Idaho Department of Correction scoresheet was originally obtained by the Electronic Privacy Information Center. *Documents Obtained by EPIC Show Idaho’s Use of Subjective Categories in Calculating Risk*, ELEC. PRIVACY INFO. CTR. (Dec. 11, 2019), <https://perma.cc/9M83-54XN>.
  4. See equivant, Practitioner’s Guide to COMPAS Core 31 (2019), <https://perma.cc/V3M9-5KGQ> [hereinafter COMPAS Guide] (stating the Pretrial Release Risk and General Recidivism scale outputs).
  5. See Alexander M. Holsinger et al., *A Rejoinder to Dressel and Farid: New Study Finds Computer Algorithm Is More Accurate than Humans at Predicting Arrest and as Good as a Group of 20 Lay Experts*, FED. PROB., Sept. 2018, at 50, 51 (“The use of actuarial assessment tools is found at every stage of the court and correctional system.”).
  6. Memorandum from the Ctr. on Sent’g & Corrs., Vera Inst. of Just. to Delaware Just. Reinvestment Task Force 4 (Oct. 12, 2011), <https://perma.cc/358K-74TW>; Thomas H. Cohen, Christopher T. Lowenkamp & William E. Hicks, *Revalidating the Federal Pretrial Risk Assessment Instrument (PTRA): A Research Summary*, FED. PROB., Sept. 2018, at 23, 23 (describing the federal government’s Pretrial Risk Assessment Instrument).

The academic literature on risk assessment instruments is voluminous but tends to center on a small cluster of issues; namely, racial parity, predictive validity, and transparency.<sup>7</sup> That focus is too narrow. Several widely used risk assessment instruments incorporate attitude and associational factors by scoring subjects' abstract beliefs, peer associations, and family criminal history.<sup>8</sup> The Level of Service Inventory–Revised (LSI-R), for example, scores subjects on whether they have an “[u]nfavorable attitude toward convention,” “[s]ome criminal acquaintances,” and “[c]riminal family/spouse.”<sup>9</sup> Such items raise significant constitutional, ethical, and empirical issues.

Evidence is not necessarily admissible just because it is useful. Courts are prohibited from considering certain information when the value of punishing crime is outweighed by the protection and vindication of a defendant's constitutional rights.<sup>10</sup> Risk assessment instruments that try to quantify attitudes and associations can run afoul of constitutionally protected freedoms. In the sentencing context, certain speech and associational evidence is subject to an absolute bar under the First Amendment.<sup>11</sup> At all stages of criminal proceedings, evidence of family criminal history should be prohibited as a matter of Fourteenth Amendment equal protection.<sup>12</sup>

Omitting attitude and associational factors from risk assessment is not just a matter of constitutionality but of pragmatism. The predictive utility of all attitude and associational factors has been overblown.<sup>13</sup> They are rife with measurement problems and bear only a nominal demonstrated correlation with recidivism.<sup>14</sup> These shortcomings bear on constitutional balancing tests

---

7. See, e.g., *The Good, Bad and Ugly of New Risk-Assessment Tech in Criminal Justice*, ABA (Feb. 16, 2020), <https://perma.cc/Q9A3-CHDB> (summarizing an expert panel discussion that raised issues of transparency and predictive accuracy); Melissa Hamilton, *Risk-Needs Assessment: Constitutional and Ethical Challenges*, 52 AM. CRIM. L. REV. 231, 257 (2015) (discussing constitutional analyses of racial issues in risk assessment).

8. See *infra* Part I.

9. Level of Service Inventory–Revised (IDOC), *supra* note 3.

10. See, e.g., *Weeks v. United States*, 232 U.S. 383, 398 (1914) (establishing a federal exclusionary rule for evidence obtained in violation of the Fourth Amendment); *Mapp v. Ohio*, 367 U.S. 643, 655 (1961) (extending the Fourth Amendment exclusionary rule to state courts); *Miranda v. Arizona*, 384 U.S. 436, 478–79 (1966) (establishing an exclusionary rule for statements made by defendants who are not properly advised of—and have not validly waived—their Fifth Amendment privilege against self-incrimination); *Crawford v. Washington*, 541 U.S. 36, 68 (2004) (holding that testimonial hearsay is only admissible under the Sixth Amendment Confrontation Clause if the declarant is unavailable and there was prior opportunity for cross-examination).

11. See *infra* Part III.A.

12. See *infra* Part IV.B.3.

13. See *infra* Parts II.B–C, IV.D.

14. See *infra* Parts II.B–C, IV.D.

for speech and family affiliation factors and counsel against the use of peer association factors.<sup>15</sup>

The few legal scholars who have addressed attitude and associational factors in criminal risk assessment have neither questioned their empirical justification nor considered that they may violate the First Amendment. Dawindar Sidhu has challenged the use of speech and associational factors; however, his challenge relies on normative arguments and reference to Tocqueville, not the First Amendment.<sup>16</sup> Melissa Hamilton has briefly addressed the constitutionality of “procriminal attitudes” as a criminal risk factor,<sup>17</sup> but only through an equal protection lens.<sup>18</sup> J.C. Oleson touches on the issue of freedom of association in an article on constitutional issues in evidence-based sentencing,<sup>19</sup> but treats it primarily as a philosophical, rather than legal, issue.<sup>20</sup> He summarily assumes that a First Amendment speech challenge could not stand: “Most courts would not uphold defendants’ challenges to evidence-based sentencing based on free speech, double jeopardy, or trial by jury rights, but some courts would be sympathetic to equal protection claims.”<sup>21</sup> Similarly, Sonja Starr challenges whether sentencing based on a defendant’s “demographic, socioeconomic, and family characteristics can be constitutionally or normatively justified” but seems to presume that

---

15. Peer association factors probably do not entail robust constitutional protections. See *infra* Part IV.

16. Dawindar S. Sidhu, *Moneyball Sentencing*, 56 B.C. L. REV. 671, 720-21 (2015).

17. Hamilton defines “procriminal attitudes” as “[a] person’s mindset towards antisocial causes.” Hamilton, *supra* note 7, at 247. For a discussion of “criminal attitude” factors in risk assessment, see Part III.B.1 below.

18. Hamilton, *supra* note 7, at 247 (“[T]hough there is no evident case law directly on point, there likely is even less concern from an equal protection standpoint of the likelihood a court would rule unconstitutional the use of factors that adjudge procriminal attitudes.”).

19. See Erin Collins, *Punishing Risk*, 107 GEO. L.J. 57, 59 (2018) (“Actuarial sentencing, also called ‘evidence-based sentencing,’ is the practice of using actuarial risk assessment tools to guide sentencing decisions.”).

20. J.C. Oleson, *Risk in Sentencing: Constitutionally Suspect Variables and Evidence-Based Sentencing*, 64 SMU L. REV. 1329, 1388-89 (2011) (“The seventeen variables associated with adult recidivism in Gendreau’s meta-analysis may prove to be philosophically problematic when employed in evidence-based sentencing. Some of those seventeen variables (e.g., antisocial personality, criminal companions, substance abuse, and even employment) are bourgeois and paternalistic in nature.”) (footnote omitted). Oleson does allude to the First Amendment right of assembly to support a normative concern about “punishing a defendant for merely associating with the ‘wrong sort of person,’” but performs no legal analysis. *Id.* at 1389.

21. *Id.* at 1376.

sentencing based on a defendant's "conduct, mental states, and attitudes" would be a justifiable alternative.<sup>22</sup>

Prior scholarship has likely overlooked problems with attitude and associational factors because it has focused on *what* factors are assessed instead of *how* factors are assessed. Identifying the issues with attitude and associational factors requires thorough examination of instrument questionnaires and scoring guides, as well as attention to the factors' empirical underpinnings. There are devils in the details. This Article aims to root them out.

## B. Article Overview

This Article is intended to be useful and accessible to all stakeholders in criminal risk assessment, including scholars, legal professionals, policymakers, and instrument developers. Legal scholars are often unfamiliar with instruments' content and empirical limitations, while other stakeholders are likely unfamiliar with the legal precepts that render certain information inadmissible in court. The legal and empirical issues in criminal risk assessment are intertwined, though. All stakeholders need the foundational legal and empirical knowledge to ensure the legal system is just and evenhanded.

Part I of this Article provides an overview of several risk assessment instruments currently used in the United States. Part II explains and critiques how risk assessment instruments quantify "criminogenic needs." Part III discusses freedom of speech and the ethical problems inherent to quantification of attitudes. Part IV addresses peer and family associational factors, delving into the constitutional quagmire that is freedom of intimate association, the reasons "criminal family" factors violate equal protection, the traces of eugenics that persist in risk assessment literature, and the ethical and statistical problems with efforts to tabulate criminal associates and family members. Finally, Part V briefly expands the scope outward to caution against implementing poorly designed and validated instruments just because risk assessment is trendy.

## I. Overview of Select Risk Assessment Instruments

This Part will provide an overview of four risk assessment instruments used by numerous American states, based on information and resources made

---

22. Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803, 819 (2014) ("[T]his Article's central question is about discrimination and disparity: whether risk prediction instruments that classify defendants by demographic, socioeconomic, and family characteristics can be constitutionally or normatively justified. One could, after all, predict risk in other ways—for instance, based only on past or present criminal behavior, or based on individual assessment of a defendant's conduct, mental states, and attitudes.").

public by developers, non-profits, scholars, and reporters. The instruments discussed in this Article were selected because they include attitude and associational factors, are not offense-specific, and are employed at sentencing in at least one state. They represent a significant fraction of the vast alphabet soup of instruments the federal government, states, and private developers have produced for use in criminal proceedings.<sup>23</sup>

Discussing the current content of risk assessment instruments requires a certain amount of guesswork and extrapolation. Criminal risk assessment instruments typically enjoy limited transparency despite their ubiquity,<sup>24</sup> and presentation of their results can vary from state to state or even county to county.<sup>25</sup> This Article presumes, based on publicly available presentence reports, briefs, and opinions, that the risk assessment report that goes to a judge includes, at most, the defendant's overall risk score, scores in each risk and need category, and a few details noted by the assessor.<sup>26</sup> The Article also

- 
23. Harvard's Berkman Klein Center for Internet & Society has assembled a database of information on more than forty criminal risk assessment instruments. *Risk Assessment Tool Database*, BERKMAN KLEIN CTR., <https://perma.cc/42C5-57NT> (archived Apr. 16, 2023).
24. See Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343, 1368-71 (2018) (citing instances in which developers invoked trade secret protections for their risk assessment instruments); Alyssa M. Carlson, Note, *The Need for Transparency in the Age of Predictive Sentencing Algorithms*, 103 IOWA L. REV. 303, 321-24 (2017) (critiquing how poorly validated instruments can be shielded from scrutiny by the invocation of trade secret protections); Cynthia Rudin, Caroline Wang & Beau Coker, *The Age of Secrecy and Unfairness in Recidivism Prediction*, 2 HARV. DATA SCI. REV., no. 1, at 4 (2020), <https://perma.cc/YG6T-EDX2> ("A separate issue with COMPAS is that it is *proprietary*, which means its calculations cannot be double-checked for individual cases, and its methodology cannot be verified.").
25. See JENNIFER K. ELEK, ROGER K. WARREN & PAMELA M. CASEY, NAT'L CTR. FOR STATE CTS., USING RISK AND NEEDS ASSESSMENT INFORMATION AT SENTENCING: OBSERVATIONS FROM TEN JURISDICTIONS 48-50 (2015), <https://perma.cc/G2JQ-Z9PF>.
26. The presentence report as a whole is far more comprehensive, but risk assessment reports are typically brief. See, e.g., *State v. Guise*, No. 17-0589, 2018 WL 2084846, at \*4 (Iowa Ct. App. May 2, 2018), *vacated*, 921 N.W.2d 26 (Iowa 2018) ("[O]ur record contains no information on what the [Iowa Risk Revised assessment] was intended to measure, how it was scored, what factors were considered in arriving at a score, or how the PSI evaluator applied the test to Guise. The IRR as described in Guise's PSI report was a black box, devoid of transparency." (citation omitted)); Presentence Investigation at 5-7, *State v. Rios*, No. CR2015-128600-001-DT (Ariz. Super. Ct. Nov. 5, 2015) [hereinafter *Rios* Presentence Investigation] (OST report); Supplemental Memorandum for Writ of Habeas Corpus app. tab 1, pt. 2 74-79, *Barnes v. Wenerowicz*, 280 F.R.D. 206 (E.D. Pa. 2012), ECF No. 18-3 [hereinafter *Barnes* Memorandum] (LSI-R report); ELEK ET AL., *supra* note 25, app. at A-121 (ORAS report).



presumes that judges do not see instruments' underlying questionnaires, interview guides, or scoring guides.<sup>27</sup>

A. Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)

COMPAS is an instrument developed in the mid-1990s by psychometrician Tim Brennan and corrections professional Dave Wells and licensed by the company equivalent (formerly Northpointe) to many jurisdictions.<sup>28</sup> COMPAS is likely the best-known and most-scrutinized criminal risk assessment instrument currently in use because of a 2016 constitutional challenge in Wisconsin and a ProPublica investigation that alleged racially biased scoring.<sup>29</sup> COMPAS scores have recently been used at sentencing in Wisconsin, California, and Michigan,<sup>30</sup> and have been used in other stages of criminal proceedings by many additional states.<sup>31</sup>

COMPAS's algorithm is a proprietary black box but its categories and some of its inputs are known.<sup>32</sup> A 2011 COMPAS questionnaire includes 137

- 
27. See, e.g., *People v. Boak*, No. 340201, 2021 WL 2290819, at \*2 (Mich. Ct. App. June 3, 2021) (recounting that the sentencing judge did not know what factors informed the risk score).
28. Eugenie Jackson & Christina Mendoza, *Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not*, 2 HARV. DATA SCI. REV., no. 1, at 5 (2020), <https://perma.cc/ZX65-MPG4>; Melissa Hamilton, *Debating Algorithmic Fairness*, 52 U.C. DAVIS L. REV. ONLINE 261, 268 n.30 (2019) ("Northpointe rebranded with the trade name equivalent . . . in January 2017."); CASEY ET AL., *supra* note 2, app. at A-20 (naming several state corrections agencies that utilize COMPAS).
29. *State v. Loomis*, 881 N.W.2d 749, 753 (Wis. 2016); Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias*, PROPUBLICA (May 23, 2016), <https://perma.cc/A4FZ-76P3>.
30. See Brief and Appendix of Defendant-Appellant at 31, *State v. Brott*, No. 2021AP002001-CR (Wis. Ct. App. Mar. 10, 2022); *People v. Howardkidd*, No. A159924, 2021 WL 1135225, at \*3 (Cal. Ct. App. Mar. 25, 2021); *Santos v. Macauley*, No. 20-cv-1226, 2021 WL 32588, at \*6 (W.D. Mich. Jan. 5, 2021).
31. See FIELD OPERATIONS ADMIN., MICH. DEP'T OF CORR., ADMINISTRATION AND USE OF COMPAS IN THE PRESENTENCE INVESTIGATION REPORT 3 n.1 (2017), <https://perma.cc/G8ZE-RTNP>; Angwin et al., *supra* note 29 (reporting pretrial use in Florida and New York); Sheldon X. Zhang, Robert E.L. Roberts & David Farabee, *An Analysis of Prisoner Reentry and Parole Risk Using COMPAS and Traditional Criminal History Measures*, 60 CRIME & DELINQ. 167, 174 (2014) (reporting parole planning use in California); RHIANA KOHL, JIQIANG RONG & HOLLIE MATTHEWS, MASS. DEP'T OF CORR., GENERAL RECIDIVISM RISK SCORE LEVELS: THREE-YEAR REVIEW ANALYSIS 1 (2018), <https://perma.cc/95JB-RV2K> (reporting inmate program assignment use in Massachusetts).
32. See *Loomis*, 881 N.W.2d at 761 ("Northpointe, Inc. . . . considers COMPAS a proprietary instrument and a trade secret. Accordingly, it does not disclose how the risk scores are determined or how the factors are weighed.").

check-box questions distributed across fifteen categories.<sup>33</sup> Because COMPAS is a proprietary product, it is not possible to ascertain how common these questions are across COMPAS instruments or whether the assessment has been updated since 2011. However, the factors discussed in the 2019 COMPAS Guide indicate that some of its products continued to include such questions, at least through 2019.<sup>34</sup>

COMPAS includes four risk scales and nineteen need scales, which are designed to be used singly or in combination at different phases of criminal proceedings.<sup>35</sup> The risk scales are intended to predict recidivism risk, while the need scales are intended to help assign individuals to appropriate interventions.<sup>36</sup> The four risk scales are Pretrial Release Risk, General Recidivism, Violent Recidivism, and Recidivism Risk Screen.<sup>37</sup> They do not appear to include any attitude or associational factors, although it is impossible to know for sure.<sup>38</sup> Among the nineteen need scales are “Criminal Associates/Peers,” “Family Criminality,” and “Criminal Thinking Self-Report.”<sup>39</sup>

---

33. Northpointe, Inc., Risk Assessment (2011), <https://perma.cc/Z5FW-PXDG> [hereinafter COMPAS Questionnaire]. The copyright date is 2011, and the Scale Set is “Wisconsin Core–Community Language.”

34. 14 of the 15 categories in the 2011 Questionnaire map directly onto 14 of the 19 “need” scales listed in the 2019 COMPAS Guide. Compare COMPAS Questionnaire, *supra* note 33, at 1-8, with COMPAS Guide, *supra* note 4, at 36-49. Of the remaining five need scales in the COMPAS Guide, four are “higher order” scales that recombine items from other need scales. See COMPAS Guide, *supra* note 4, at 36, 38, 45, 47. Although “Anger” in the questionnaire and “Financial Problems” in the guide have no definite correlates, their items could be folded into other categories (e.g., “Anger” could be folded into “Criminal Personality”). See COMPAS Questionnaire, *supra* note 32, at 7-8; COMPAS Guide, *supra* note 4, at 42, 49.

35. COMPAS Guide, *supra* note 4, at 2, 30-32, 36-50. COMPAS has forty-three scales in total, including seventeen specific to women, but the Guide names and discusses only four risk scales, nineteen need scales, and a “Lie Scale and Random Responding Test.” See *id.* at 2, 49-50.

36. *Id.* at 21-22 (“The needs scales are not meant to be predictive but aim simply and accurately to describe the offender along dimensions relevant for correctional practice.”).

37. *Id.* at 31-32.

38. *Id.* states that the “risk scales make limited use of dynamic variables.” *Id.* at 12. Dynamic variables are factors that a subject can change, like attitudes or associations. See *infra* Part IIA.1. The Pretrial Risk and Recidivism Risk Screen scales do not include attitude or associational factors. See COMPAS Guide, *supra* note 4, at 31-33. The COMPAS Practitioner’s Guide uses fuzzy language to obscure exactly what inputs factor into the General Recidivism and Violent Recidivism scales, but none of the stated factors relate to attitudes or associations. See *id.* at 31-32 (listing several factors that the two scales “include”); *Risk Scores: The Not-So-Secret Recipe*, EQUIVANT (Aug. 14, 2020), <https://perma.cc/4WN9-7A96> (listing the same factors as the COMPAS Guide plus arrest rate as risk score elements).

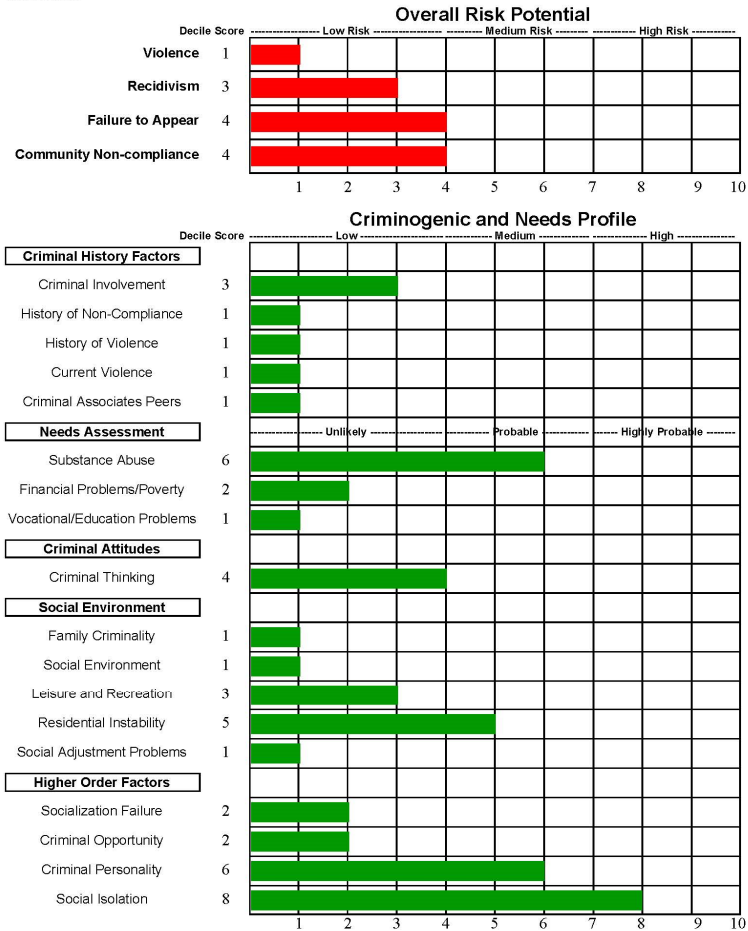
39. COMPAS Guide, *supra* note 4, at 36, 39.

COMPAS score reports are, in some instances, structured to intimate that need factors are significant predictors of recidivism, as evidenced by a sample COMPAS risk assessment report secured by ProPublica and reproduced here. The report does not disclose that need scales are poor predictors of recidivism, as Part II.B will discuss.

**Figure 1**  
Sample COMPAS Score Report

**Northpointe COMPAS Risk Assessment**

Name: **Class3, Jessie**                      SSN:                      Offender #: **01cr57**  
 Date of Birth: **06/19/1977**              Date of Screening: **08/14/2006**  
 Comment:



The report is entitled “Northpointe COMPAS Risk Assessment” in bold letters at the top—“Risk Assessment,” not “Risk and Needs Assessment.” It does not designate any needs in the “Criminogenic and Needs Profile” as noncriminogenic. The most intuitive reading of this chart—based on the title, the probability language in the “Criminogenic and Needs Profile” section (“Unlikely”/“Probable”/“Highly Probable”), and the absence of interpretive instructions—is that all the factors are substantially correlated with recidivism risk. Additionally, the contrast between the generality of the risk scale titles (e.g., “Recidivism”) and the specificity of the need scale titles (e.g., “Criminal Thinking” and “Criminal Personality”) would lead many readers to believe that the need scales are included in the calculation of the risk scores.

Alternate presentations of COMPAS needs assessments pose similar problems. A presentation developed by the Michigan Department of Corrections separates out “needs” from “risk,” but is still structured to imply that need factors are strongly correlated with recidivism risk.<sup>40</sup>

**Figure 2**  
Sample Needs Assessment

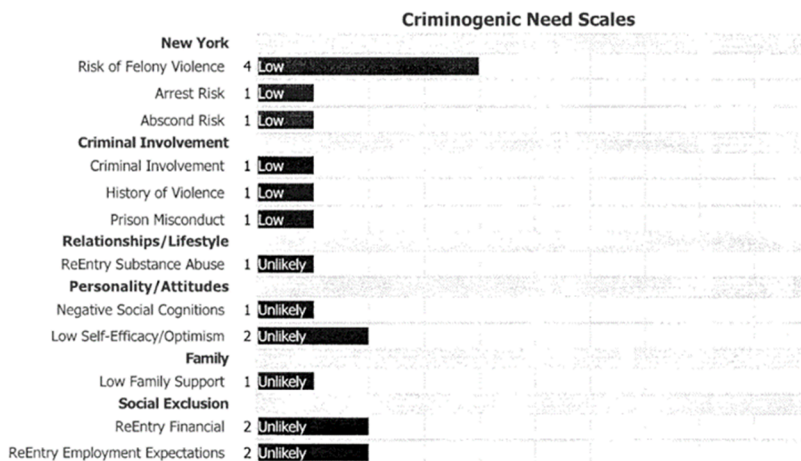
NEEDS ASSESSMENT

| Name: Beech, Adam         |                 | Number:012345   | Date:4/21/2015 |
|---------------------------|-----------------|---|----------------|
| Core COMPAS Need Scale    | Scale Score     | Supervision Recommendation  |                |
| Criminal Associates/Peers | Unlikely        | None  |                |
| Criminal Opportunity      | Probable        | Impose curfew, Structure daily activities, frequent reports                   |                |
| Leisure/Recreation        | Probable        | Assess for cognitive issues, develop prosocial supports                       |                |
| Social Isolation          | Unlikely        | None  |                |
| Substance Abuse           | Unlikely        | None  |                |
| Criminal Personality      | Highly Probable | Assess for cognitive issues   |                |
| Cognitive/Behavioral      | Probable        | Assess for cognitive issues   |                |
| Family Criminality        | Probable        | Monitor contact with family, Emphasize involvement with positive peers/family |                |
| Vocational/Education      | Highly Probable | Assess for GED and Employment Skills/Work Programs                            |                |
| Residential Instability   | Unlikely        | None  |                |
| Social Environment        | Unlikely        | None  |                |
| Other:                    | N/A             |   |                |

40. FIELD OPERATIONS ADMIN., *supra* note 31, at 32.

A sample 2020 New York score report developed by equivant designates all needs as criminogenic, uses probabilistic language, and can easily be misread to imply that the need scales inform the risk scores.<sup>41</sup>

**Figure 3**  
2020 New York Score Report



There is no question that some judges and attorneys interpret COMPAS need scales as significant predictors of recidivism. Examples from Wisconsin appellate briefs and sentencing transcripts include:

[T]he presentence investigation reported Sanders’ results on the COMPAS test to determine Sanders’ risk of recidivism. He was categorized as a low risk of violent and general recidivism, his history of violence is low, and his criminal personality and ability to control anger were ranked “unlikely” to result in recidivism.<sup>42</sup>

\* \* \*

In addition to identifying general levels of risk to re-offend, COMPAS also identifies criminogenic needs specific to that offender which are most likely to affect future criminal behavior. So when you’ve got it saying that, to me it’s pointing at . . . these needs and that is exactly what this should be used for.

41. See Fordham Parole Info. Project, COMPAS Risk Assessment-FUSL000093 (2020-01-21) (2022), <https://perma.cc/KQ4F-73GN>.

42. Appellant’s Brief at 44, State v. Sanders, Nos. 2014AP2644CR, 2014AP2645 & 2014AP2646 (Wis. Ct. App. June 2, 2016).

*What Even Is a Criminal Attitude?*  
75 STAN. L. REV. 1431 (2023)

If you have these high needs, you can use that to determine and attempt to predict criminal behavior.<sup>43</sup>

\* \* \*

In each and every [COMPAS category] that was assigned an actual risk assessment, the report writer stated that [the defendant's] risk was unlikely. Some of the specific comments include: . . .

2. "The Criminal Thinking Self-Report Scale suggests that [the defendant] is unlikely to rationalize his criminal behavior. He may exhibit remorse and regret for his behavior."<sup>44</sup>

These judges and attorneys made a rational inference that need scores are an empirically sound way to evaluate risk. The COMPAS Practitioner's Guide admits that they are not,<sup>45</sup> but there is no reason to think judges read the guide.

#### B. Level of Service Inventory–Revised (LSI-R)

The LSI-R is a commercial instrument originally developed for use with Canadian probationers<sup>46</sup> by criminologist Donald Andrews and clinical psychologist James Bonta.<sup>47</sup> It is intended to "[m]atch the level of service to the level of risk" by prioritizing higher-risk offenders for "treatment" and matching subjects to appropriate treatments based on their needs and personal characteristics.<sup>48</sup> The LSI-R questionnaire includes fifty-four scored, equally weighted items across ten categories.<sup>49</sup> LSI-R scores are based on a semi-

---

43. *State v. Samsa*, 859 N.W.2d 149, 153-54 (Wis. Ct. App. 2014).

44. Brief of Defendant-Appellant at \*7, *State v. Graf*, No. 2018AP001081, 2019 WL 13186716 (Wis. Ct. App. Sept. 17, 2018) (citation omitted), 2018 WI APP. CT. BRIEFS LEXIS 937.

45. See COMPAS Guide, *supra* note 4, at 21 (explaining that need scales "are not meant to be predictive").

46. JAMES BONTA, SUMMARY OF RESEARCH FINDINGS ON THE LSI (1993), in *LSI-R: LEVEL OF SERVICE INVENTORY-REVISED: INTRODUCTORY TRAINING COURSE—PARTICIPANT MANUAL* 31, 31 (n.d.), <https://perma.cc/5ZLR-2GE4>.

47. CASEY ET AL., *supra* note 2, app. at A-31 (identifying the LSI-R's original developers). See generally JAMES BONTA & D.A. ANDREWS, *THE PSYCHOLOGY OF CRIMINAL CONDUCT* (6th ed. 2017) (identifying Bonta as a psychologist and researcher with a Ph.D. in clinical psychology and Andrews as a criminologist).

48. BONTA & ANDREWS, *supra* note 47, at 191-92.

49. BONTA, *supra* note 46, at 32 ("The items of the LSI are scores fairly simple [sic]. There are no fancy weighting formulas that run the risk of making computational errors. Further, research comparing simple scoring methods like the 0-1 format with more statistically complex methods show [sic] the simple approach is as valid as the more complex procedures.").

structured interview, in which the interviewer and subject engage in conversation with open-ended questions.<sup>50</sup>

The LSI-R has been billed by its company as “[t]he most widely used and researched risk/need assessment in the world.”<sup>51</sup> It certainly appears to be the most widely used risk assessment instrument in state presentence reports: LSI-R scores have been considered at sentencing in Alabama, Illinois, Kansas, and Oklahoma in recent years.<sup>52</sup> LSI-R scores have also been considered at sentencing in Hawaii, Idaho, Indiana, Maine, Maryland, Minnesota, North Carolina, North Dakota, Ohio, Texas, Utah, and Wisconsin.<sup>53</sup> Many U.S. states have used the LSI-R at other points in the criminal justice process, as well.<sup>54</sup>

LSI-R reports include a risk score, recommendations for institutional classification, a bar chart showing level of “Risk/Needs” in each category on a five-point scale, selected details on “subcomponent risk/needs,” and a summary of responses to the fifty-four items.<sup>55</sup>

---

50. LEVEL OF SERVICE INVENTORY (LSI-R) TRAINING WORKSHOP: PARTICIPANT’S GUIDE 33-36 (2017), <https://perma.cc/UE96-H43M>.

51. *LSI-R: Level of Service Inventory-Revised*, MULTI-HEALTH SYS., <https://perma.cc/AD7E-K233>, (archived June 23, 2023) (to locate, select “View the live page”); see also *LSI-R: Level of Service Inventory-Revised*, MULTI-HEALTH SYS., <https://perma.cc/Z3B8-SW2Q> (archived Apr. 19, 2023) (billing the LSI-R as “[t]he predecessor to the LS/CMI, the most widely used and researched risk/need assessment”); CASEY ET AL., *supra* note 2, app. at A-31 (“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.”).

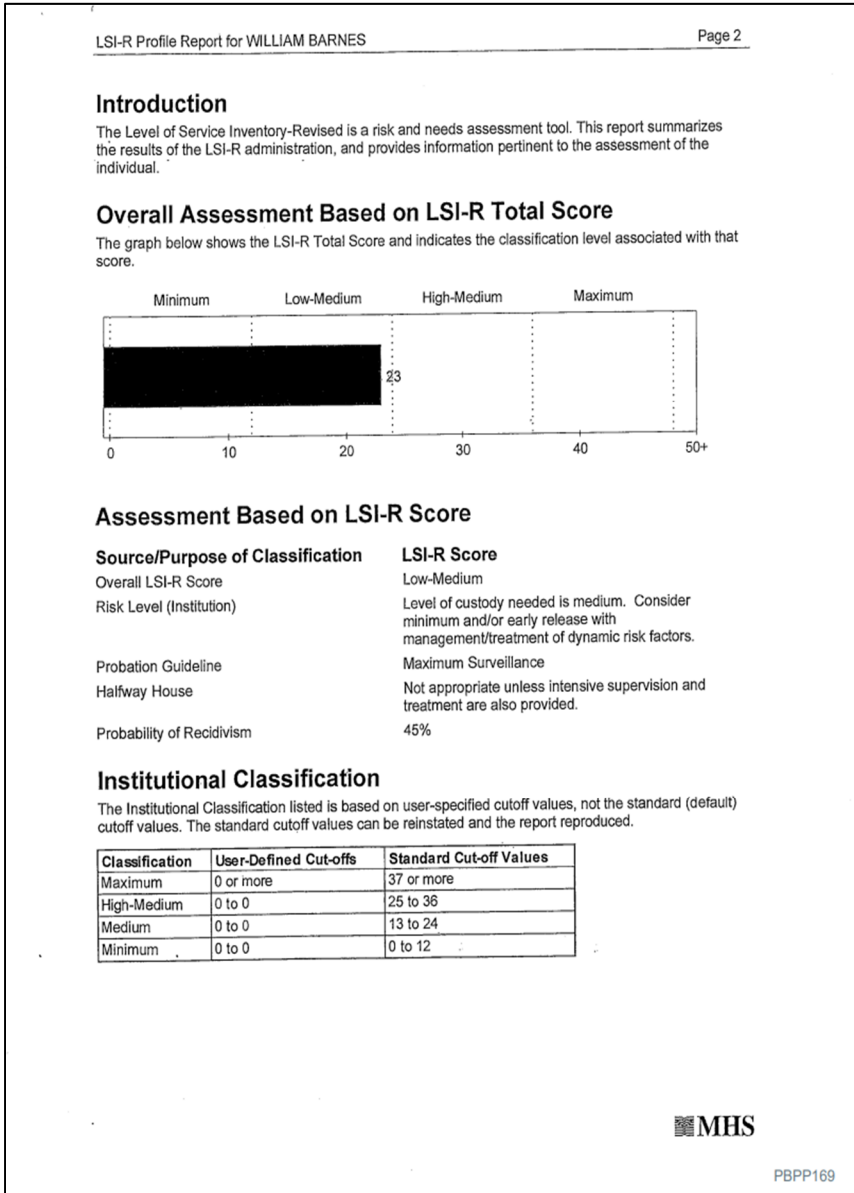
52. See *United States v. Gamble*, No. 07-cr-00219, 2018 WL 3812241, at \*3 (M.D. Ala. Aug. 10, 2018); *People v. Ashley*, 2021 IL 180688-U, ¶ 36 (App. Ct. 2021); *State v. Baska*, No. 121739, 2020 WL 4380944, at \*8 (Kan. Ct. App. July 31, 2020); *State v. Jones*, No. CF-16-5816, 2018 Okla. Dist. LEXIS 3080, at \*5-6 (Feb. 15, 2018).

53. *State v. Gauthier*, 939 A.2d 77, 81 (Me. 2007); *State v. Barnhart*, No. ot-10-032, 2011 WL 5317301, at \*2 (Ohio Ct. App. Nov. 4, 2011); *Bonnee v. State*, No. 14-11-00603-cr, 2012 WL 3862029, at \*2 (Tex. Ct. App. Sept. 6, 2012); ELEK ET AL., *supra* note 25, app. at A-26 (Colorado); *id.* app. at A-40 (Idaho); *id.* app. at A-148 (Wisconsin); Tammy Howell, Multi-Health Sys., LSI-R, LS/RNR and LS/CMI Documentation (n.d.), <https://perma.cc/4HQT-2NPX> (Hawaii, Maryland, Minnesota, North Carolina, North Dakota, and Utah).

54. See Howell, *supra* note 53 (reporting usage for probation and/or parole in Connecticut, Indiana, New Jersey, Pennsylvania, and Utah).

55. See MULTI-HEALTH SYS., LSI-R PROFILE REPORT: REX DARLINGTON 3, 5 (2001), <https://perma.cc/8X27-J6VU> (to locate, select “View the live page”); Barnes Memorandum, *supra* note 26, at 74-79.

**Figure 4<sup>56</sup>**  
2011 Pennsylvania Assessment



56. Barnes Memorandum, *supra* note 26, at 74-79.



*What Even Is a Criminal Attitude?*  
75 STAN. L. REV. 1431 (2023)

**Assessment of Risk/Needs Based on LSI-R Subcomponents**

The graph below displays specific areas, and indicates whether they are low, medium, or high risk/needs areas.



**Details Regarding Subcomponent Risks/Needs**

**Criminal History**

1. Any prior adult convictions: Yes
2. Two or more prior convictions: Yes
3. Three or more prior convictions: Yes
4. Three or more present offenses: Yes
5. Arrested under age 16: Yes
6. Ever incarcerated upon conviction: Yes
7. Escape history from a correctional facility: Yes
8. Ever punished for institutional misconduct: Yes, Number: 1
9. Charge laid or probation/parole suspended during prior community supervision: Yes
10. Official record of assault/violence: Yes

**Education/Employment**

11. Currently unemployed: Yes
12. Frequently unemployed: Yes
15. Less than regular grade 10: Yes
16. Less than regular grade 12: Yes
18. Participation/performance: A very unsatisfactory situation with a very clear and strong need for improvement
19. Peer interactions: A very unsatisfactory situation with a very clear and strong need for improvement
20. Authority interactions: A very unsatisfactory situation with a very clear and strong need for improvement



**Financial**

22. Reliance upon social assistance: Yes

**Family/Marital**

24. Non-rewarding, parental: A very unsatisfactory situation with a very clear and strong need for improvement

26. Criminal-Family/Spouse: Yes

**Companions**

33. Some criminal acquaintances: Yes

34. Some criminal friends: Yes

**Attitudes/Orientation**

51. Supportive of crime: A relatively unsatisfactory situation with a need for improvement

What Even Is a Criminal Attitude?  
75 STAN. L. REV. 1431 (2023)

LSI-R Profile Report for WILLIAM BARNES

Page 6

### Summary of LSI-R Item Responses

The rater entered the following response values for the items on the Level of Service Inventory-Revised Form.

| Item | Response | Item | Response | Item | Response |
|------|----------|------|----------|------|----------|
| 1.   | Y        | 19.  | 0        | 37.  | N        |
| 2.   | Y        | 20.  | 0        | 38.  | N        |
| 3.   | Y        | 21.  | 2        | 39.  | 2        |
| 4.   | Y        | 22.  | Y        | 40.  | 2        |
| 5.   | Y        | 23.  | 2        | 41.  | N        |
| 6.   | Y        | 24.  | 0        | 42.  | N        |
| 7.   | Y        | 25.  | 2        | 43.  | N        |
| 8.   | Y        | 26.  | Y        | 44.  | N        |
| 9.   | Y        | 27.  | 2        | 45.  | N        |
| 10.  | Y        | 28.  | N        | 46.  | N        |
| 11.  | Y        | 29.  | N        | 47.  | N        |
| 12.  | Y        | 30.  | N        | 48.  | N        |
| 13.  | N        | 31.  | 2        | 49.  | N        |
| 14.  | N        | 32.  | N        | 50.  | N        |
| 15.  | Y        | 33.  | Y        | 51.  | 1        |
| 16.  | Y        | 34.  | Y        | 52.  | 2        |
| 17.  | N        | 35.  | N        | 53.  | N        |
| 18.  | 0        | 36.  | N        | 54.  | N        |

#### Additional Item Information

- 1. Number of prior convictions: Not Specified
- 4. Number of present offenses: Not Specified
- 8. Number of times punished for institutional misconduct: 1
- 40. Type of drug associated with current drug problem: 1
- 45. Other indicators of drug problem: None Specified
- 50. Area of psychological assessment indicated: None Specified

Date Printed: Friday, February 04, 2011

End of Report

 MHS

PBPP173

The LSI-R’s reliance on semi-structured interviews has a drawback: Although open-ended questions allow for more nuance and flexibility than closed-ended questions, they also increase the likelihood that assessor bias or inconsistencies distort the results. For example, Idaho’s LSI-R scoring guide offers the following distinction between a score of 2 and a score of 3 on the “Attitudes/Orientation—Unfavorable Toward Convention” item:<sup>57</sup>

**Table 1**  
LSI-R Scoring Guide

| 2  | 3   |
|--|---|
| The offender is supportive of a relatively pro-social lifestyle. Has a few ties to conventional settings (home, family, school, work) that are fairly strong. Is generally rejecting of criminal values and orientation, but may benefit from some further values clarification. | The offender is committed and invested in society’s institutions (government, business, family, school, spiritual institutions) and their underlying values (order, peace, justice, love, etc.). The offender indicates positive functions of conventional activities, e.g., working, studying, playing sports, etc., and places an emphasis on rewards of such activities. |

It is hard to see the daylight between those two descriptions; assessor discretion will play an enormous role in drawing the distinction. In fact, a 2003 LSI-R revalidation study sponsored by the Pennsylvania Board of Probation and Parole found a 72% agreement rate among assessors on that item, which fell below the Board’s “minimally acceptable performance standard” of 80%.<sup>58</sup> In that example, and in similar items throughout the LSI-R, the integer risk score lends assessor discretion a veneer of objectivity, accuracy, and

---

57. Idaho LSI-R Scoring Guide—Version 3.0, at 26-27 (2015), <https://perma.cc/QUS5-VX56>. The “2” column in the scoring guide also includes a note, omitted above for conciseness, that “[t]he offender’s unsolicited endorsements of conventional norms, and behavior throughout the interview are as important in scoring as any responses to specific probes in this area.” *Id.* at 27.

58. JAMES AUSTIN, DANA COLEMAN, JOHNETTE PEYTON & KELLY DEDEL JOHNSON, INST. ON CRIME, JUST. & CORRECTIONS, GEORGE WASHINGTON UNIV., RELIABILITY AND VALIDITY STUDY OF THE LSI-R RISK ASSESSMENT INSTRUMENT 12, 14 (2003). This study was performed twelve years prior to the publication of the Idaho LSI-R Scoring Guide. See Idaho LSI-R Scoring Guide—Version 3.0, *supra* note 57. However, the study’s text suggests a similar scoring process: “Through an interview process,” the authors wrote, “offenders are rated on items requiring either a ‘yes/no’ response, or the use of a structured scale ranging in value from 0 to 3.” AUSTIN ET AL., *supra*, at 6.

mathematical precision.<sup>59</sup> Consequently, a decisionmaker who is alert to potential bias or omissions in a narrative presentence report may be less alert to the possibility of bias in LSI-R risk scores.

### C. Ohio Risk Assessment System (ORAS) and Its Derivatives

ORAS is a set of risk assessment instruments developed by the University of Cincinnati on behalf of the Ohio Department of Rehabilitation and Correction.<sup>60</sup> The set includes the Community Supervision Tool (“ORAS-CST”), which is meant to assist probation decisionmaking.<sup>61</sup> ORAS-CST assessments are designed to be administered through a structured interview, a self-report questionnaire, and corroboration by official records and collateral sources.<sup>62</sup>

ORAS tools are used throughout the state of Ohio and in many other jurisdictions at multiple stages of criminal proceedings. Ohio law requires the Department of Rehabilitation and Correction to provide an ORAS report to sentencing judges upon request,<sup>63</sup> and some Ohio counties explicitly require presentence investigation reports to include ORAS results.<sup>64</sup> The Ohio Judicial Conference has endorsed the use of ORAS-CST at sentencing<sup>65</sup> and it appears

---

59. See Andrea Nishi, Note, *Privatizing Sentencing: A Delegation Framework for Recidivism Risk Assessment*, 119 COLUM. L. REV. 1671, 1685-86 (2019) (discussing the “false sense of objectivity that surrounds algorithmic predictions”).

60. ELEK ET AL., *supra* note 25, app. at A-104 n.3.

61. EDWARD LATESSA, PAULA SMITH, RICHARD LEMKE, MATTHEW MAKARIOS & CHRISTOPHER LOWENKAMP, UNIV. OF CINCINNATI SCH. OF CRIM. JUST., CREATION AND VALIDATION OF THE OHIO RISK ASSESSMENT SYSTEM: FINAL REPORT 22 (2009), <https://perma.cc/8GN8-ZST7> (“The CST is designed to assist in both designation of supervision level, as well as to guide case management for offenders in the community.”).

62. UNIV. OF CINCINNATI CTR. FOR CRIM. JUST. RSCH., OHIO RISK ASSESSMENT SYSTEM—COMMUNITY SUPERVISION TOOL (ORAS-CST) INTERVIEW GUIDE, in 2 NEV. JUD. COUNCIL, COMM. TO STUDY EVIDENCE-BASED PRETRIAL RELEASE, THE NEED FOR REFORM 2-22, 2-36 to -38 (n.d.), <https://perma.cc/7UYK-N78S>.

63. OHIO REV. CODE ANN. § 5120.114(A) (LexisNexis 2019) (requiring the department to provide an assessment using a “single validated risk assessment tool” at sentencing judges’ request); OHIO ADMIN. CODE 5120-13-1(B) (2023) (designating ORAS as the single risk assessment tool).

64. See, e.g., *Probation Department*, CHAMPAIGN CNTY. CT. COMMON PLEAS, <https://perma.cc/27YM-49W2> (archived Apr. 19, 2023) (“The Champaign County Common Pleas Court, General Division, is required to use the ORAS evaluation tool when considering criminal sentences.”); ELEK ET AL., *supra* note 25, app. at A-107 (stating that all Cuyahoga County presentence reports include ORAS results).

65. Cmty. Corr. Comm., Ohio Jud. Conf., Policy Statement on the Ohio Risk Assessment System and Risk and Needs Assessment Tools 1 (2015), <https://perma.cc/LC4R-PSAR> (“[W]e support the use of the tool as one of several means to inform and assist judges in sentencing decisions regarding offender risk reduction and management within the community.”).

to be the norm statewide for judges to do so.<sup>66</sup> Indiana and Texas use modified versions of ORAS (IRAS and TRAS, respectively) to prepare presentence reports.<sup>67</sup> Several other states and counties also employ ORAS-derived tools in pretrial or post-sentencing decisionmaking.<sup>68</sup>

ORAS-CST and its derivatives are quite different from other instruments in several respects. One unique feature is that they merge some categories that other instruments tend to treat as distinct. “Criminal attitudes” and “criminal behavioral patterns,” for example, are collapsed into a single category, as are “associations” and “activities.”<sup>69</sup>

The ORAS-CST “criminal attitudes” section is different from those in other instruments in that the questions are mostly about the subject’s feelings and behavior and make little inquiry into abstract beliefs. ORAS-CST materials instruct the assessor to ask a series of open-ended questions:

- How do you feel about what happened?
  - What do you think about crime?
  - Tell me about the victims?
  - How do you think they feel about what you did?
- As a general rule do you worry about other people’s problems?
- Do you sometimes feel that you have lost control over events in your life? Why?
- Do you think it is sometimes ok to tell a lie? Under what circumstances?

---

66. See OHIO CRIM. SENT’G COMM’N, FELONY SENTENCING IN OHIO: THEN, NOW, AND NOW WHAT? 55 (2022), <https://perma.cc/S8P6-PX4K> (stating that “utilization of ORAS is now the common denominator in matters of policy, practice and procedure” surrounding sentencing and other aspects of the criminal justice process).

67. See IND. OFF. OF COURT SERVS., INDIANA PROBATION STANDARDS 11 (2020), <https://perma.cc/9F2E-YLZ8> (requiring Indiana probation departments to use IRAS results to “provide the court with information regarding risk, need and responsivity factors” and to make recommendations concerning supervision, sanctions, programs, and services); ELEK ET AL., *supra* note 25, app. at A-135 (stating that Texas officials adopted the Texas Risk Assessment System as a statewide standard).

68. See, e.g., *Ohio Risk Assessment System (ORAS)*, MO. DEP’T CORR., <https://perma.cc/TCN3-YEA8> (archived Apr. 19, 2023); N.H. DEP’T OF CORR., REINTEGRATION STUDY: PART 3: TRANSITIONAL PLANNING & SUPPORT 2 (2018), <https://perma.cc/L7G9-YKBQ>; *Pretrial Release: Risk Assessment Tools (SB 36)*, CAL. CTS., <https://perma.cc/B4MH-Y32K> (archived Apr. 19, 2023); *Statewide Collaborative Offender Risk Evaluation System (SCORES)*, STATE CONN. BD. PARDONS & PAROLES, <https://perma.cc/EK6U-7S8P> (archived Apr. 19, 2023); Jacob Holmes, *State Senator Aims to Emphasize Risk Assessment in Parole Considerations*, ALA. POL. REP. (Oct. 27, 2021, 7:25 AM CDT), <https://perma.cc/Q242-PZJ4>.

69. UNIV. OF CINCINNATI CTR. FOR CRIM. JUST. RSCH, *supra* note 62, at 2-22, 2-36 to 2-38.

*What Even Is a Criminal Attitude?*  
75 STAN. L. REV. 1431 (2023)

- Do you consider yourself to be a risk taker?
  - How about when you committed your offense?
  - How did it make you feel?
  - If I asked you to rate yourself as a risk taker on a scale from 1 to 5 with 1 being not at all, and 5 being often how would you rate yourself?
- Would you describe yourself as someone who “Walks away from a fight”, or “Tries to avoid it but it seems to find you” or, “first one in”?
  - If I asked you to rate yourself on a scale from 1 to 5 with 1 being “walks away”, and 5 being “first one in” how would you rate yourself?
- Have you ever heard the saying, “Do Unto Others Before They Do Unto You”? In general, do you:
  - Disagree with statement,
  - Depends on the situation, or
  - Agree with statement
- How do you feel about getting some help or participating in programs?<sup>70</sup>

The assessor assigns a score of 0 to 2 in seven items based on the subject’s answers: “Criminal Pride,” “Expresses Concern about Other’s [sic] Misfortunes,” “Feels Lack of Control over Events,” “Sees No Problem in Telling Lies,” “Engages in Risk Taking Behavior,” “Walks Away from a Fight,” and “Believes in ‘Do Unto Others Before They Do Unto You.’”<sup>71</sup>

Another unique characteristic of ORAS-CST is that it includes a separate, unscored “responsivity” assessment, which identifies “special considerations or responsivity factors that might affect the offender’s engagement in supervision or programming.”<sup>72</sup> The assessor completes a checklist of twelve factors including physical disability, low motivation, mental health issues, limited English, and history of abuse/neglect.<sup>73</sup> Several of those items are scored factors in other assessments, but ORAS-CST does not score the responsivity

---

70. *Id.* at 2-36 to -38.

71. OHIO RISK ASSESSMENT COMMUNITY SUPERVISION TOOL, in MEGAN E. COLLINS, EMILY M. GLAZENER, CHRISTINA D. STEWART & JAMES P. LYNCH, DEPT. OF CRIMINOLOGY & CRIM. JUST., U. MD., FOLLOW-UP REPORT TO THE MSCCSP: USING ASSESSMENT INSTRUMENTS DURING CRIMINAL SENTENCING app. E at 60 (2015), <https://perma.cc/D5XW-YVTT>. The TRAS-CST interview guide includes a substantively identical set of questions. U. CINCINNATI & TEX. DEP’T CRIM. JUST., *Texas Risk Assessment System Community Supervision Interview Guide—Felony*, in TEXAS RISK ASSESSMENT SYSTEM PARTICIPANT MANUAL 24, 37-39 (2014); UNIV. OF CINCINNATI & TEX. DEP’T OF CRIM. JUST., *Texas Risk Assessment System, Community Supervision Interview Guide—Misdemeanor*, in TEXAS RISK ASSESSMENT SYSTEM PARTICIPANT MANUAL, *supra*, at 56, 67-68.

72. UNIV. OF CINCINNATI CTR. FOR CRIM. JUST. RSCH., *supra* note 62, at 2-39.

73. *Id.*

factors.<sup>74</sup> They are not intended to predict risk but to identify potential barriers to “treatment.”<sup>75</sup>

D. Offender Screening Tool (OST) and Modified Offender Screening Tool (MOST)

1. OST and MOST content—known and inferred

The OST is a forty-four-item instrument developed by the probation department of Maricopa County, Arizona, and forensic psychologist David Simourd.<sup>76</sup> It has been used statewide in Arizona presentence reports for probation-eligible cases since 2005.<sup>77</sup> It is also used by Virginia probation departments in misdemeanor cases.<sup>78</sup> The Arizona OST includes forty-two scored items across nine domains.<sup>79</sup> Two additional items in a tenth domain, “Physical Health/Medical,” appear to be scored in the Virginia OST but not the Arizona OST.<sup>80</sup>

---

74. LATESSA ET AL., *supra* note 61, at 18 (“Responsivity items are not used in the final calculation of risk, but instead are used as case planning factors that should be addressed to improve [the] likelihood that programming will reduce recidivism.”).

75. *Id.* (explaining that the developers included a responsivity assessment in ORAS to gather information on items that “are not directly related to recidivism, but instead have the potential to restrict the efficacy of treatment”).

76. COLLINS ET AL., *supra* note 71, at 11; *Offender Screening Tool (OST)*, ARIZ. JUD. BRANCH, <https://perma.cc/NN2C-G65R> (archived Apr. 19, 2023).

77. ARIZ. CODE JUD. ADMIN. § 6-201.01 (2020) (“For all probation-eligible cases, presentence reports shall also contain case information related to criminogenic risk and needs as documented by the standardized risk assessment and other file and collateral information.”); Order Adopting the Standardized Assessment and Reassessment Tool and Conducting a Pilot Program for Reassessment Timeframes for Adult Intensive Probationers, Arizona Sup. Ct. Admin. Order No. 2005-12 (2005), <https://perma.cc/J69A-YYF7> (“For purposes of establishing a statewide standardized assessment tool, the Administrative Office of the Courts reviewed, selected, and validated the Offender Screening Tool (OST) and Field Reassessment Offender Screening Tool (FROST).”).

78. CASEY ET AL., *supra* note 2, app. at A-45.

79. *Offender Screening Tool (OST)*, *supra* note 76.

80. See NAT’L CTR. FOR STATE CTS., VIRGINIA MOST/OST REVALIDATION AND PROXY RISK NORMING: RECIDIVISM RESULTS FY12, at 23 (2016) (including “Physical Health/Medical” as a scored domain); 127 Offender Screening Tool (OST) Graph, in *Rios Presentence Investigation*, *supra* note 26, at app. (no “Physical Health/Medical” domain); COCONINO CNTY. ADULT PROB. DEP’T, PRESENTENCE REPORT-OFFENDER SCREENING TOOL RESULTS, in ELEK ET AL., *supra* note 25, app. at A-8, app. at A-8 to A-9 (noting that “[t]he Physical/Medical Health category has been identified as a responsivity factor and is not shown to impact risk level”).



The MOST is a condensed, eight-item version of the OST, used for defendants “processed through an expedited or early disposition court.”<sup>81</sup> Subjects who receive a high score on the MOST in Arizona are administered the full OST.<sup>82</sup> OST and MOST scores are tabulated using scoresheets, likely based on answers to standardized questionnaires.<sup>83</sup>

There is limited publicly available information on the OST and MOST, but one can make plausible inferences about the current content from what is available. The most comprehensive source of information is a 2002 report by Penny Stinson, then-director of pretrial services for Maricopa County,<sup>84</sup> which includes appendices with OST and MOST materials.<sup>85</sup> A more recent source of information is a 2016 revalidation study of Virginia’s OST and MOST, which includes an appendix with OST and MOST scoresheets.<sup>86</sup> Although those are the only two sources for OST and MOST background materials, OST score reports as recent as 2016 are available on Lexis as an element of Arizona presentence reports.<sup>87</sup>

Each of Arizona’s fifteen counties makes an independent decision on how to display OST scores.<sup>88</sup> Maricopa County, which includes more than half of Arizona residents, and Coconino County have used an information-rich format that includes the defendant’s overall risk score and level, score in each domain, and bullet points in each domain providing further narrative detail.<sup>89</sup>

---

81. ELEK ET AL., *supra* note 25, app. at A-4 n.3; PENNY STINSON, INST. FOR CT. MGMT., DEVELOPMENT OF A RISK ASSESSMENT TOOL TO BE USED IN BAIL RELEASE DECISIONS IN MARICOPA COUNTY, ARIZONA 106 (2002).

82. ELEK ET AL., *supra* note 25, app. at A-4 n.3.

83. *See* STINSON, *supra* note 81, app. 3 at 98-102, app. 4 at 106.

84. *Id.* at i.

85. *Id.* app. 3 at 80-103 (OST screening tool); *id.* app. 4 at 104-06 (MOST screening tool).

86. NAT’L CTR. FOR STATE CTS., *supra* note 80, app. A at 20-25; *id.* app. B at 26.

87. *See, e.g.*, Presentence Investigation at 2-4, *State v. Dunlop*, No. CR2014-148799-001-DT (Ariz. Super. Ct. Oct. 9, 2015); Revised Presentence Report—Part One at 7-13, *Arizona v. Frodsham*, No. CR 201600521 (Ariz. Super. Ct. Nov. 7, 2016), *reprinted in* Exhibit A at 32, *Frodsham ex rel. Fleming & Curti PLC v. Arizona*, No. 22-cv-01074, 2022 WL 3082911 (D. Ariz. June 23, 2022).

88. ELEK ET AL., *supra* note 25, app. at A-4 to -5.

89. *See, e.g.*, *Rios Presentence Investigation*, *supra* note 26, at 5-7; Coconino Cnty. Adult Prob. Dep’t, Presentence Report-Offender Screening Tool Results: State of Arizona v. Diaz, Arthur Kevin (n.d.), in JUD. COUNCIL OF CAL., CALIFORNIA RISK ASSESSMENT PILOT PROJECT TRAINING 64, 64-65 (2010), <https://perma.cc/WB86-GQPY>. A version of the Diaz presentence report that includes the scores can be found at Coconino Cnty. Prob. Dep’t, Arizona Report 1, in PAMELA M. CASEY, ROGER K. WARREN & JENNIFER K. ELEK, NAT’L CTR. FOR STATE CTS., USING OFFENDER RISK & NEEDS ASSESSMENT INFORMATION AT SENTENCING: GUIDANCE FOR COURTS FROM A NATIONAL WORKING GROUP app. B at 2, 2-7 (2015), <https://perma.cc/3TLP-4GAK>.

**Figure 5<sup>90</sup>**  
Maricopa County OST Report

| The Superior Court of Arizona in Maricopa County – Adult Probation Department  |                   |                     |  |
|--|-------------------|---------------------|--|
| Chief Probation Officer Barbara A. Broderick   |                   |                     |  |
| State of Arizona v. Anthony Rios, II CR2014-153351-001-DT, CR2015-116610-001-DT, and CR2015-128600-001-DT  |                   |                     |  |
| <b>Interested Party Statements:</b>  |                   |                     |  |
| The Maricopa Deputy County Attorney has provided a sentencing recommendation, which is attached for the Court.   |                   |                     |  |
| <b>Risk/Needs Assessment:</b>  |                   |                     |  |
| The Offender Screening Tool (OST) uses a detailed interview and scoring system to determine an offender’s risk to re-offend and identifies the risk factors (domains) of their life that would benefit from intervention. There are nine domains grouped by their impact on the offender’s overall risk level. Scores of 60% or higher (50% for mental health domain) in each individual domain contribute the most to the offender’s total risk level, whereas lower scores may contribute moderately or not at all. The identified risk factors provide the Court and other agencies information to match supervision and treatment interventions that will address an offender’s greatest risk factors, potentially reducing their risk to re-offend. The OST has been validated for statewide use on Arizona’s adult population and has been approved by the Administrative Office of the Courts. A graph showing this defendant’s scores on the OST is attached to this report. |                   |                     |  |
| <b>OVERALL RISK/NEEDS LEVEL</b><br>(Male Risk/Needs Range)   |                   |                     |  |
|  |                   |                     | <b>26</b>  |
| Low (0-5)  | Medium-Low (6-10) | Medium-High (11-17) | High (18-42)   |
| DOMAINS THAT DO NOT CONTRIBUTE TO OVERALL RISK/NEED LEVEL  |                   |                     |  |
| DOMAIN   | SCORE             |                     | SUMMARY  |
| Education  | 0/3               | 0%                  | <ul style="list-style-type: none"> <li>• 10<sup>th</sup> grade education</li> <li>• Dropped out to help his mother</li> <li>• Interested in obtaining his G.E.D.</li> <li>• Would like to further his education</li> </ul>                           |
| DOMAINS THAT MAY CONTRIBUTE TO OVERALL RISK/NEED LEVEL   |                   |                     |  |
| DOMAIN   | SCORE             |                     | SUMMARY  |
| Family and Social Relationships  | 4/8               | 50%                 | <ul style="list-style-type: none"> <li>• Primarily raised by an Aunt</li> <li>• Father was in prison while he was growing up</li> <li>• Mother spent most of her time with “boyfriends”</li> <li>• Moved out of the family home at age 16</li> </ul> |
| PAGE 5   |                   |                     |  |

90. Rios Presentence Investigation, *supra* note 26, at 5-7.

*What Even Is a Criminal Attitude?*  
75 STAN. L. REV. 1431 (2023)

**The Superior Court of Arizona in Maricopa County – Adult Probation Department**  
*Chief Probation Officer Barbara A. Broderick*

State of Arizona v. Anthony Rios, II CR2014-153351-001-DT, CR2015-116610-001-DT, and  
CR2015-128600-001-DT

|             |     |     |   |
|-------------|-----|-----|---|
|             |     |     | <ul style="list-style-type: none"> <li>• Described his childhood as "not so good"</li> <li>• Claims to have a significant other</li> <li>• Peers are considered to be negative and criminally oriented</li> </ul>   |
| Residential | 1/2 | 50% | <ul style="list-style-type: none"> <li>• Native to Arizona</li> <li>• Has had periods of living a homeless lifestyle</li> </ul>   |
| Attitude    | 3/7 | 43% | <ul style="list-style-type: none"> <li>• Expressed attitudes that would be considered non-conforming to societal norms</li> <li>• Feels he needs to change his lifestyle and does need help to change</li> <li>• Willing to comply with community supervision</li> <li>• Claims he is motivated to change his life</li> <li>• Did not report any gang involvement</li> <li>• Per the police report, the defendant is a documented gang member of "Westside Phoeniquera 27<sup>th</sup> Avenue" street gang</li> </ul> |

**DOMAINS THAT SIGNIFICANTLY CONTRIBUTE TO OVERALL RISK/NEED LEVEL**

| DOMAIN            | SCORE |      | SUMMARY   |
|-------------------|-------|------|---|
| Drug Abuse        | 3/3   | 100% | <ul style="list-style-type: none"> <li>• Began using illegal drugs at the age of 9</li> <li>• Has used and/or abused marijuana, cocaine, methamphetamine, and prescription drugs</li> <li>• Prior to being incarcerated he was using methamphetamine and marijuana</li> <li>• Drug abuse has caused him legal, financial, and health problems</li> <li>• Feels he does have a drug problem and in need of treatment</li> <li>• Has previously participated in 2 substance abuse treatment programs</li> </ul> |
| Criminal Behavior | 8/9   | 89%  | <ul style="list-style-type: none"> <li>• 2 prior felony convictions, 1 for violent offense</li> <li>• 13 prior misdemeanor convictions, 4 for violent offenses</li> <li>• Previous grants of standard probation</li> <li>• Previous probation revocations</li> <li>• Previous terms of prison</li> </ul>  |
| Vocational        | 4/5   | 80%  | <ul style="list-style-type: none"> <li>• Currently unemployed</li> </ul>  |

PAGE 6

*What Even Is a Criminal Attitude?*  
75 STAN. L. REV. 1431 (2023)

**The Superior Court of Arizona in Maricopa County – Adult Probation Department**  
*Chief Probation Officer Barbara A. Broderick*

State of Arizona v. Anthony Rios, II CR2014-153351-001-DT, CR2015-116610-001-DT, and  
CR2015-128600-001-DT

|             |     |     |   |
|-------------|-----|-----|---|
|             |     |     | <ul style="list-style-type: none"> <li>• Described his childhood as “not so good”</li> <li>• Claims to have a significant other</li> <li>• Peers are considered to be negative and criminally oriented</li> </ul>   |
| Residential | 1/2 | 50% | <ul style="list-style-type: none"> <li>• Native to Arizona</li> <li>• Has had periods of living a homeless lifestyle</li> </ul>   |
| Attitude    | 3/7 | 43% | <ul style="list-style-type: none"> <li>• Expressed attitudes that would be considered non-conforming to societal norms</li> <li>• Feels he needs to change his lifestyle and does need help to change</li> <li>• Willing to comply with community supervision</li> <li>• Claims he is motivated to change his life</li> <li>• Did not report any gang involvement</li> <li>• Per the police report, the defendant is a documented gang member of “Westside Phoenix 27<sup>th</sup> Avenue” street gang</li> </ul> |

DOMAINS THAT SIGNIFICANTLY CONTRIBUTE TO OVERALL RISK/NEED LEVEL

| DOMAIN            | SCORE |      | SUMMARY   |
|-------------------|-------|------|---|
| Drug Abuse        | 3/3   | 100% | <ul style="list-style-type: none"> <li>• Began using illegal drugs at the age of 9</li> <li>• Has used and/or abused marijuana, cocaine, methamphetamine, and prescription drugs</li> <li>• Prior to being incarcerated he was using methamphetamine and marijuana</li> <li>• Drug abuse has caused him legal, financial, and health problems</li> <li>• Feels he does have a drug problem and in need of treatment</li> <li>• Has previously participated in 2 substance abuse treatment programs</li> </ul> |
| Criminal Behavior | 8/9   | 89%  | <ul style="list-style-type: none"> <li>• 2 prior felony convictions, 1 for violent offense</li> <li>• 13 prior misdemeanor convictions, 4 for violent offenses</li> <li>• Previous grants of standard probation</li> <li>• Previous probation revocations</li> <li>• Previous terms of prison</li> </ul>  |
| Vocational        | 4/5   | 80%  | <ul style="list-style-type: none"> <li>• Currently unemployed</li> </ul>  |

PAGE 6

*What Even Is a Criminal Attitude?*  
75 STAN. L. REV. 1431 (2023)

**The Superior Court of Arizona in Maricopa County – Adult Probation Department**  
*Chief Probation Officer Barbara A. Broderick*

State of Arizona v. **Anthony Rios, II** CR2014-153351-001-DT, CR2015-116610-001-DT, and  
CR2015-128600-001-DT

|               |     |     |  |
|---------------|-----|-----|--|
|               |     |     | <ul style="list-style-type: none"><li>• Last worked doing telemarketing</li><li>• Family provided additional financial support</li><li>• Has difficulty meeting his monthly financial obligations</li></ul>  |
| Alcohol Use   | 2/3 | 67% | <ul style="list-style-type: none"><li>• First drank age 15</li><li>• Was drinking occasionally to intoxication</li><li>• Alcohol abuse has caused him his legal problems</li><li>• Feels he does have an alcohol problem and could benefit from counseling</li><li>• Longest period of sobriety occurred while serving a prison sentence</li></ul> |
| Mental Health | 1/2 | 50% | <ul style="list-style-type: none"><li>• See confidential section of report</li></ul>   |

**Additional Assessments and Screenings:**

The defendant is eligible for intensive probation supervision.

**Financial Assessment:**

The defendant is currently unemployed due to his incarcerated status. He last worked in telemarketing and received additional financial support from his family. The defendant does appear capable of obtaining and maintaining gainful employment. There has been no request for restitution at this time. It is recommended the Court retain jurisdiction over the matter of restitution for any future requests. The defendant should be accountable for any other Court-ordered financial assessments. Reimbursement for legal fees will not be recommended as the Court has found the defendant to be indigent.

**Discussion and Evaluation:**

The defendant is presently before the Court for sentencing for his third, fourth, fifth, and sixth felony convictions. The defendant committed numerous offenses over a period of approximately of eight months and also committed the offenses while he was on parole or felony release. The defendant would like to receive a minimum prison sentence and feels he was not represented appropriately.

Standard assessments completed on the defendant score him as a high risk to reoffend. The domains of vocation, alcohol abuse, drug abuse, and criminal behavior do contribute significantly to his risk. The domains that might contribute to his risk include social/family relations, residence, mental health, and attitude. He appears to have an adequate education that could be furthered.

Cochise County has used an even more information-rich format that separates out risk, need, and protective factors in each domain, as well as a format that does not include any scores.<sup>91</sup> Mohave County has used a format that groups narrative information by OST domain but includes only the overall risk/need score, not subsidiary domain scores.<sup>92</sup> Yavapai County has utilized a similar format but with a non-numerical overall risk level.<sup>93</sup> Presentence reports from the other ten counties are not available on Lexis, Westlaw, or any public web sources.<sup>94</sup>

Understanding the constitutional and normative implications of the OST and MOST requires review of the instrument questionnaires, not just the scoresheets. The questionnaire items are far more troubling than the scoresheet items. For example, OST scoresheet item (b)—“Does client have attitudes that are non-conforming to social norms”—appears to correspond to OST questionnaire items (d), (e), and (f)—“I think it is okay to have tattoos or

---

91. See Revised Presentence Report—Part One, *supra* note 87, at 7-9; Revised Presentence Report—Part Two at 10-13, *Arizona v. Frodsham*, No. CR 201600521 (Ariz. Super. Ct. Nov. 7, 2016), reprinted in Exhibit A at 32, *Frodsham ex rel. Fleming & Curti PLC v. Arizona*, No. 22-cv-01074, 2022 WL 3082911 (D. Ariz. June 23, 2022).

92. See Presentence Investigation Report at 5-6, *State v. Gomez-Torres*, No. CR-2013-01013 (Ariz. Super. Ct. Oct. 20, 2014).

93. See Presentence Investigation Report at 4-6, *Arizona v. Richards*, No. P1300CR201600476 (Ariz. Super. Ct. May 16, 2022).

94. I did not find any Arizona presentence reports in Westlaw. I performed Google searches for information on probation offices and presentence reports in each Arizona county for which I found no presentence reports in Lexis. No county included a sample presentence report or, in most instances, disclosed that probation officers would use a risk assessment tool. See *Probation Services*, APACHE CNTY., <https://perma.cc/DLG8-YARS> (archived Apr. 19, 2023) (no presentence information); *Felony Procedures*, GILA CNTY., <https://perma.cc/7DBN-RBXZ> (archived Apr. 19, 2023) (“The Pre-sentence Report will discuss the defendant’s life and any other crimes s/he may have committed, and will contain a recommendation for a specific sentence.”); *Probation*, GRAHAM CNTY., <https://perma.cc/AU7T-2Q65> (archived Apr. 19, 2023) (no presentence information); *Departments & Services*, GREENLEE CNTY., <https://perma.cc/X7L9-V6ET> (archived Apr. 19, 2023) (no probation information at all); *About Probation*, LA PAZ CNTY., <https://perma.cc/APN3-DQ87> (archived Feb. 7, 2023) (disclosing the use of risk assessment only for sex offenders); La Paz County Probation-Pre-Sentence Intake Questionnaire (n.d.) (including a Presentence Interview Packet that does not contain an OST assessment); *Probation*, NAVAJO CNTY., <https://perma.cc/USK9-GRLF> (archived Apr. 19, 2023) (stating that the department assesses needs and risk but not discussing the presentence report or OST); *Pre-Sentence Process*, PIMA CNTY., <https://perma.cc/H68R-32R5> (archived May. 15, 2023) (noting assessments are completed in the presentence report interview and investigation process); *Presentence Investigations*, PINAL CO., <https://perma.cc/Q8RG-D6RZ> (archived Apr. 19, 2023) (no disclosure of risk assessment); *Adult Probation Services Division (APSD)*, SANTA CRUZ CNTY., <https://perma.cc/462F-Q52Z> (archived June 10, 2023) (no disclosure of presentence risk assessment); *Presentence Investigations*, YUMA CNTY., <https://perma.cc/3MN4-MJGN> (archived Apr. 19, 2023) (no disclosure of risk assessment).

bodypiercing,” “I believe that society is to blame for many problems in the world,” and “Sometimes a person has to do whatever it takes to make money, even if [it] is not exactly what would be considered ‘normal.’”<sup>95</sup>

**Table 2**  
OST/MOST “Attitude” Items and Questions

| <b>OST Scoresheet “Attitude” Items<br/>(Arizona 1999 and Virginia 2016)<sup>96</sup></b>  |
|---|
| <ul style="list-style-type: none"> <li>a) Does client have attitudes supportive of crime?</li> <li>b) Does client have attitudes that are non-conforming to societal norms?</li> <li>c) Does client have a poor attitude about his/her current [conviction/offense]?</li> <li>d) Does client have a poor attitude about community supervision?</li> <li>e) Does client have a poor attitude toward authority figures?</li> <li>f) According to the Screener, what is the client’s motivation level to improve his/her life?</li> <li>g) According to the Screener, this client’s need for improvement in attitude is:</li> </ul> <p>[Items (a) to (e) are yes/no questions. Item (f) responses are scored 0 for “Good to Fair” and 1 for “Poor.” Item (g) responses are scored 0 for “None to Low” and 1 for “Moderate to High.”]</p> |

| <b>OST “Attitude” Questionnaire<br/>(Arizona 1999)</b>   |
|--|
| <ul style="list-style-type: none"> <li>a) For the most part, I think that it is okay to commit crime.</li> <li>b) I believe that most people would agree there is easy money to be made in crime.</li> <li>c) I think most people would commit crime if they could get away with it.</li> <li>d) I think it is okay to have tattoos or bodypiercing.</li> <li>e) I believe that society is to blame for many problems in the world.</li> <li>f) Sometimes a person has to do whatever it takes to make money, even if [it] is not exactly what would be considered ‘normal.’</li> <li>g) What do you think about being convicted on your current offense?</li> <li>h) Do you plan to appeal your conviction?</li> <li>i) My current conviction makes me really think hard about trying to change my life.</li> <li>j) I think I may need some help if I am to change my life.</li> <li>k) Will you be upset if you receive some type of community supervision (e.g., probation, parole, or FARE) for your current conviction?</li> </ul> |

*table continued on next page*

---

95. See STINSON, *supra* note 81, at 93, 101.

96. *Id.* at 101; NAT’L CTR. FOR STATE CTS., *supra* note 80, at 25.

*What Even Is a Criminal Attitude?*  
75 STAN. L. REV. 1431 (2023)

| <b>OST "Attitude" Questionnaire<br/>(Arizona 1999)—Continued</b>  |
|---|
| <p>l) Do you think you will have a problem complying with conditions of some type of community supervision (e.g., probation, parole, or FARE)?</p> <p>m) Do you think the police really help anybody?</p> <p>n) Do you think school teachers just like to have power over the students?</p> <p>o) Do you think work bosses like to order people to do things?</p> <p>p) Have you ever been a member or associate of a gang?</p> <p>q) Are you still active in a gang?</p> |

| <b>VA MOST Scoresheet<br/>"Attitude" Item<br/>(Virginia 2016)<sup>97</sup></b>   | <b>AZ MOST Scoresheet<br/>"Attitude" Item<br/>(Arizona 1999)<sup>98</sup></b>   | <b>AZ MOST "Attitude"<br/>Questionnaire<br/>(Arizona 1999)<sup>99</sup></b>   |
|--|---|---|
| <p><i>To assess if the client has attitudes that are non-conforming to societal norms</i></p> <p>Tell me about your present offense? Your criminal history? What are your thoughts about breaking the law?</p> <p>(Listen for: rationalizations and minimizations about criminal behavior, oppositional, defiance with authority...)</p> <p>[Score of 0 for "No, Prosocial" and 1 for "Yes, Antisocial." There is a line below the item for comments.]</p> | <p>Does client have attitudes that are non-conforming to societal norms?</p> <p>[Score is 0 for "No" and 1 for "Yes." No space for comment is provided]</p> | <p>a) I think it is okay to have tattoos or bodypiercing.</p> <p>b) I believe that society is to blame for many problems in the world.</p> <p>c) Sometimes a person has to do whatever it takes to make money, even if it is not exactly what would be considered "normal."</p> <p>[A score of 1 on any of these questions will result in a score of 1 on M-OST item F ["Attitude"].]</p> |

---

97. NAT'L CTR. FOR STATE CTS., *supra* note 80, at 21.

98. STINSON, *supra* note 81, at 106.

99. *Id.* at 105.



It is hard to understand why the designers selected only OST items (d), (e), and (f) for the MOST “Attitude” domain when other OST “Attitude” items concern stances on crime, personal accountability, and personal behavior, not abstract beliefs on social and moral questions.<sup>100</sup> The MOST “Attitude” scoresheet assigns one point (out of eight possible total points) to any subject who agrees with any one of the three items. The scoresheets do not reveal that MOST subjects score a point if they do not condemn tattoos or body piercing; the OST and MOST’s design flaws emerge in full only when the questionnaires are examined.

It is impossible for an outsider to ascertain whether the questionnaire attached to the 2002 Stinson report, or something similar, is still in use, but two things are clear: (1) At least at the beginning, the questionnaire was standardized and required, and (2) the questionnaire was still in use as of 2015. The Stinson report asserted that “[t]he OST has been incorporated into the presentence questionnaire to provide a seamless process of gathering offender information.”<sup>101</sup> A 2002 probation department report noted that OST assessors used software that locked them into answering every single question that contributed to the score, to ensure uniformity in administration.<sup>102</sup>

A 2012 Maricopa County presentence report shows quite obviously that the questionnaire was still in use at that time, as the “Attitude” summary quotes directly from the questionnaire.<sup>103</sup>

---

100. *See id.* at 93-94, 105.

101. *Id.* at 40.

102. Jennifer L. Ferguson, *Putting the “What Works” Research into Practice: An Organizational Perspective*, 29 CRIM. JUST. & BEHAV. 472, 487 (2002).

103. Presentence Investigation at 3, *State v. Martinez*, No. CR2010-006085-001-DT (Ariz. Super. Ct. Apr. 6, 2012).

**Table 3**<sup>104</sup>  
OST “Attitude” Questionnaire—Select Items

| OST Questionnaire Items Echoed in 2012 “Attitude” Assessment |  |
|--|--|
|  | <ul style="list-style-type: none"> <li>• I believe that most people would agree there is easy money to be made in crime.</li> <li>• I think most people would commit crime if they could get away with it.</li> <li>• What do you think about being convicted on your current offense?</li> <li>• My current conviction makes me really think hard about trying to change my life.</li> <li>• Have you ever been a member or associate of a gang?</li> </ul> |

**Figure 6**<sup>105</sup>  
2012 OST “Attitude” Assessment

| DOMAIN   | SCORE      | SUMMARY   |
|----------|------------|---|
| Attitude | 4/7<br>57% | <ul style="list-style-type: none"> <li>• The defendant feels there is easy money to be made in crime.</li> <li>• He feels most people would commit crime if they could get away with it.</li> <li>• He feels his conviction is fair, and he is motivated to make changes in his life.</li> <li>• He has been a member of a criminal street gang.</li> </ul> |

Maricopa County presentence reports from 2013 to 2015 also intimate that the 1999 questionnaire was still in use.<sup>106</sup> The OST questionnaire double-dips on criminal associates, including both a distinct “Social Relationships” domain and a gang affiliation item in the “Attitude” domain.<sup>107</sup> Other instruments include gang affiliation only in their social relationships category.<sup>108</sup> The “Attitude” domain on the OST scoresheet, as opposed to the questionnaire, does not inquire into gang affiliation at all on its face;<sup>109</sup> however, Maricopa County presentence reports from 2013 to 2015 include notes in the “Attitude” summary about gang membership.<sup>110</sup>

104. STINSON, *supra* note 81, at 93-94.

105. *Id.*; Presentence Investigation, *supra* note 104, at 93-94.

106. *See, e.g.*, Presentence Investigation, *supra* note 112, at 3; Rios Presentence Investigation, *supra* note 26, at 6.

107. STINSON, *supra* note 81, at 94, 99.

108. ORAS has a “Gang Membership” item in the “Peer Associations” category. COLLINS ET AL., *supra* note 71, at 60. COMPAS asks several questions about gang membership in the “Peers” category. COMPAS Questionnaire, *supra* note 33, at 3. The LSI-R does not explicitly inquire into gang membership.

109. *See* STINSON, *supra* note 81, at 101.

110. *See, e.g.*, Presentence Investigation, *supra* note 112, at 3; Rios Presentence Investigation, *supra* note 26, at 6.

**Figure 7<sup>111</sup>**  
OST Scoresheet Excerpts, 2013 & 2015

|          |     |     |   |
|----------|-----|-----|---|
| Attitude | 6/7 | 86% | <ul style="list-style-type: none"> <li>• The defendant may have attitudes that are supportive of crime.</li> <li>• The defendant may have attitudes that are non-conforming to societal norms.</li> <li>• The defendant believes he was mistreated in the present offense.</li> <li>• He has a history of failure to comply with probation conditions.</li> <li>• He has expressed little motivation to change his life.</li> <li>• He was a member of West Side City gang from age twelve to twenty-four.</li> </ul> |
| Attitude | 3/7 | 43% | <ul style="list-style-type: none"> <li>• Expressed attitudes that would be considered non-conforming to societal norms</li> <li>• Feels he needs to change his lifestyle and does need help to change</li> <li>• Willing to comply with community supervision</li> <li>• Claims he is motivated to change his life</li> <li>• Did not report any gang involvement</li> <li>• Per the police report, the defendant is a documented gang member of “Westside Phoeniquera 27<sup>th</sup> Avenue” street gang</li> </ul> |

That inclusion suggests that the probation officers relied on the 1999 questionnaire, or something close to it, to interview the subjects about their attitudes.

## 2. OST and MOST design problems

Irrespective of whether the original questionnaires are still in use, the OST and MOST are the most poorly designed and constitutionally dubious instruments discussed in this Article. Both instruments include items that, by the admission of their developers, bear only a weak correlation with recidivism.<sup>112</sup> The Stinson report describes three groups of items selected for inclusion in the OST and MOST, in descending order of “statistical connection within the current data.”<sup>113</sup>

---

111. See, e.g., Presentence Investigation at 3, *State v. Johnson*, No. CR2011-114400-003-DT (Ariz. Super. Ct. Aug. 2, 2013); *Rios* Presentence Investigation, *supra* note 26, at 6.

112. See STINSON, *supra* note 81, at 44 (describing the inclusion of items with a “modest” or “generally weaker statistical connection” to available data (quoting David Simourd, Development of the Mini Offender Screening Tool (MOST), Presentation to the Am. Probation & Parole Ass’n Conf. (Aug. 2001))).

113. *Id.* at 43-44.

- 1) “[S]ome items (e.g., prior felony, and prior adjudications) were included because they had the greatest statistical connection within the current data.”<sup>114</sup>
- 2) “Others (e.g., current drug effects, number of alcohol arrests, employment history) were included because they had [a] modest statistical connection within the current data and were supported by the general correctional research literature on their relevance in criminogenic assessment.”<sup>115</sup>
- 3) “Finally, some items (e.g., partner, peers[,] non-conformist attitude) were included because they had great practical utility in spite of their generally weaker statistical connection within the current data.”<sup>116</sup>

To be clear, the developers of the OST and MOST admitted, at the time the instruments were developed and first implemented, that attitude and associational items had a “generally weaker statistical connection” to recidivism, which is evidently a lower bar than even a “modest statistical connection.”<sup>117</sup> The actual numbers are a mystery. The Stinson report does not share any OST/MOST validation data or analysis, and neither of the two Arizona OST validation studies (performed in 2003 and 2008) is publicly available.<sup>118</sup>

The OST and MOST do not assign lesser weight to their less-correlated factors. The MOST assigns a score of 1 for “Attitude” if the subject agrees with any of the three “Attitude” items.<sup>119</sup> Consequently, a subject’s opinion on whether it is “okay to have tattoos and bodypiercing” receives the same weight in their MOST risk score as a prior felony conviction or current unemployment.<sup>120</sup> The OST includes seven “Attitude” items but only two items for “Residence and Neighborhood” and three items apiece for “Alcohol” and “Drug Abuse.”<sup>121</sup> Since each item is worth one point, the subject’s “attitudes that are non-conforming to societal norms” receive the same weight as more intuitively criminogenic items like “[h]istory of problems with drug use” and “[d]oes the offender have any previous felony convictions?”<sup>122</sup>

Compounding the OST’s construction problems, Arizona presentence reports classify domains’ relevance to “risk/need” based on the defendant’s

---

114. *Id.*

115. *Id.* at 44.

116. *Id.* (quoting Simourd, *supra* note 113, at 1-17).

117. *See id.*

118. *See* CASEY ET AL., *supra* note 2, app. at A-47 (stating that the two OST validation studies occurred).

119. STINSON, *supra* note 81, at 105.

120. *See id.* at 104-05.

121. *Id.* at 100-01.

122. *See id.* at 100-02.

score in each domain, not on the domain’s statistical correlation with recidivism. Each of the nine domains is grouped into one of three categories:

- 1) “DOMAINS THAT DO NOT CONTRIBUTE TO OVERALL RISK/NEED LEVEL”
- 2) “DOMAINS THAT MAY CONTRIBUTE TO OVERALL RISK/NEED LEVEL”
- 3) “DOMAINS THAT SIGNIFICANTLY CONTRIBUTE TO RISK/NEED LEVEL”<sup>123</sup>

If a defendant receives a 0/7 score on the “Attitude” domain, then that domain is designated as one that does not contribute to overall risk/needs level.<sup>124</sup> But if the defendant receives a 1/7, then “Attitude” is designated a domain that “may” contribute to overall risk/needs level.<sup>125</sup>

**Figure 8**<sup>126</sup>

Arizona Presentence Report Excerpt

| DOMAINS THAT MAY CONTRIBUTE TO OVERALL RISK/NEED LEVEL |       |     |   |
|--|-------|-----|---|
| DOMAIN   | SCORE |     | SUMMARY   |
| Attitude   | 1/7   | 14% | <ul style="list-style-type: none"> <li>• Believes a conviction in the present offense is fair</li> <li>• Has attitudes non-conforming to societal norms</li> <li>• Reports wanting to make positive changes in his life prior to his most recent arrest</li> </ul>  |
| Family and Social Relationships                        | 1/8   | 13% | <ul style="list-style-type: none"> <li>• Raised by both biological parents</li> <li>• Has one sibling</li> <li>• Maintains positive, supportive relationships with his family members</li> <li>• Not currently involved in a romantic relationship</li> <li>• He has been divorced for approximately seven years</li> <li>• Has three children from two previous relationships</li> <li>• Peer associations are positive and law-abiding</li> </ul> |

A defendant who receives a score above 0 on “Attitude” will, therefore, present as having an elevated risk of recidivism because of their attitudes, even if there is little to no correlation between the OST’s “Attitude” section and

123. See, e.g., Presentence Investigation, *supra* note 87, at 3-4.

124. See, e.g., Expedited Presentence Investigation at 3-4, State v. Celaya, No. CR2010-124769-007-DT (Ariz. Super. Ct. June 10, 2013).

125. See Presentence Investigation, *supra* note 87, at 3.

126. *Id.*

recidivism. In both design and presentation, the OST overemphasizes factors weakly correlated with recidivism.

## II. How Risk-Needs Quantification Works and Doesn't Work

### A. How Risk is Measured

#### 1. Typical risk assessment factors

As a rule, instrument developers adhere to the “risk-need-responsivity (RNR) model” of offender assessment and treatment promulgated by the developers of the LSI-R, James Andrews and Daniel Bonta.<sup>127</sup> The model attempts to (1) match “intensity of service” to risk level, (2) identify “criminogenic needs,” and (3) employ appropriate, individualized strategies to address those needs.<sup>128</sup>

Instruments designed according to the RNR model generate one or more risk scores, perform a quantitative assessment of the subject’s purported needs, and identify responsivity factors that will facilitate or hinder rehabilitation.<sup>129</sup> Different types of factors inform the risk and need portions of the assessment. Factors like family criminal history, prior arrests, and prior convictions are “static”—the subject cannot change them, so they cannot be targets for intervention.<sup>130</sup> Factors like attitudes, substance abuse, and employment are “dynamic”—they can be changed and are, therefore, regarded as potential targets for intervention.<sup>131</sup> Risk factors can be static or dynamic, but need factors are necessarily dynamic.<sup>132</sup> Some instruments fold need factors into risk score calculations while others separate out risk assessment from needs assessment.<sup>133</sup> Responsivity factors may be static or dynamic and are sometimes unscored.<sup>134</sup>

---

127. See BONTA & ANDREWS, *supra* note 47, at 191-92; see also NATHAN JAMES, CONG. RSCH. SERV., R44087, RISK AND NEEDS ASSESSMENT IN THE CRIMINAL JUSTICE SYSTEM 4 (2015) (“The Risk-Needs-Responsivity (RNR) model is one of the most dominant paradigms in the risk and needs assessment field. . . . Experts in the field of risk and needs assessment assert that assessment systems should adhere to the RNR model.”).

128. BONTA & ANDREWS, *supra* note 47, at 176-77.

129. See CASEY ET AL., *supra* note 2, at 10-11.

130. *Id.* at 6.

131. *Id.*

132. *Id.*

133. *Id.* at 10 (“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.”).

134. See, e.g., PRESENTENCE REPORT-OFFENDER SCREENING TOOL RESULTS, *supra* note 80, app. at A-8 to -9 (noting that the OST “Physical/Medical Health category has been identified  
*footnote continued on next page*

Andrews and Bonta developed not only the RNR model, but also the concepts of the “big four” and “central eight” criminogenic factors, which have been accepted as canon in risk assessment literature even though they are not well-supported by empirical research.<sup>135</sup> The big four factors are “antisocial attitudes, antisocial associates, antisocial personalities, and criminal history”; they are rounded out in the central eight by “substance abuse, family characteristics, education and employment, and lack of prosocial leisure or recreation.”<sup>136</sup> Because leaders in the field promote the concepts of the “big four” and “central eight,” those concepts influence instrument development and policy recommendations.<sup>137</sup> Andrews and Bonta, however, revised their position on the “big four” in recent years, writing in 2017 that research on various populations of offenders does not support a demarcation between the big four and the other factors.<sup>138</sup> The discussion below will demonstrate that demoting the big four does not go far enough—the concept of the central eight should be dissolved altogether because dynamic risk factors have only a nominal demonstrated correlation with recidivism and are extremely hard to measure.

---

as a responsivity factor and is not shown to impact risk level”); NAT’L CTR. FOR STATE CTS., *supra* note 80, at 23 (scoring the “Physical Health/Medical” OST domain); LATESSA ET AL., *supra* note 61, at 18 (asserting that ORAS responsivity factors “such as intelligence and literacy” are used for case planning but not risk score calculations).

135. *See, e.g.*, D.A. Andrews, James Bonta & J. Stephen Wormith, *The Recent Past and Near Future of Risk and/or Need Assessment*, 52 CRIME & DELINQ. 7, 10-11 (2006); Grant Duwe, *The Development, Validity, and Reliability of the Minnesota Screening Tool Assessing Recidivism Risk (MnSTARR)*, 25 CRIM. JUST. POL’Y REV. 579, 597 (2014).

136. Hamilton, *supra* note 7, at 235-36.

137. *Id.* (“Experts maintain there exists a ‘central eight’ risk-needs categories that research consistently show are most associated with recidivism. . . . Thus, risk-needs instruments in the field of criminal offending often embed at least a few factors from the central eight categories.”); CASEY ET AL., *supra* note 2, app. at A-20 (“The COMPAS scales also include key offender risk and needs factors that have emerged from meta-analytic research, including the ‘central 8.’”); Duwe, *supra* note 136, at 597 (explaining that the MnSTARR instrument’s design is consistent with existing research, including the big four and central eight factors); CASEY ET AL., *supra* note 89, at 5 (including in its endorsement of the risk-need-responsivity model a chart asserting that “Antisocial Personality Pattern,” “Procriminal Attitudes,” and “Social Supports for Crime” are the “Major Criminogenic (Dynamic Risk) Factors” that are “Most Related to Recidivism” and that the other dynamic factors in the central eight are “Also Related to Recidivism”); FIELD OPERATIONS ADMIN., *supra* note 31, at 19 (explaining that the Michigan Department of Corrections uses COMPAS scales connected to the central eight factors for case planning).

138. BONTA & ANDREWS, *supra* note 47, at 44.

2. Commonly used statistics: AUC and  $r$

Risk assessment research and literature commonly uses two statistics, area under the receiver operating characteristic curve (AUC) and correlation coefficients ( $r$ ). These statistics can be calculated both for an instrument as a whole and subsidiary scales within an instrument. Neither AUC nor  $r$  is intuitive to a lay audience, so it is important for stakeholders to understand what they do and do not say about instruments, factors, and items.

a. AUC

AUC is a comparative measure of true positives (e.g., the number of subjects labeled “high risk” who recidivated) and false positives (e.g., the number of subjects labeled “high risk” who did not recidivate) when a predictive instrument is applied to a sample group.<sup>139</sup> In a risk assessment context, it represents the likelihood that, if you randomly select a recidivist from a sample, they will have a higher risk score than a randomly selected non-recidivist.<sup>140</sup> AUC values range from 0.50 (50% accurate, or no better than pure chance) to 1.00 (100% accurate).<sup>141</sup> Criminal risk assessment instruments generally produce AUCs around 0.70.<sup>142</sup>

To provide a clear framework for understanding AUC, consider an instrument with an AUC of 0.70 that divides subjects into two groups, high-risk and low-risk. There is a 70% chance that a recidivist will have a higher risk score on that instrument than a non-recidivist. Since this instrument only has two risk categories, if you choose a perfect cut-off score, the instrument would make the right prediction 70% of the time but fail 30% of the time.

AUC usage has important caveats related to variation within a population. Instruments typically divide subjects into more than two risk groups,<sup>143</sup> and

---

139. See CASEY ET AL., *supra* note 2, at 17; MELISSA HAMILTON, NAT’L ASS’N OF CRIM. DEF. LAWS., RISK ASSESSMENT TOOLS IN THE CRIMINAL LEGAL SYSTEM—THEORY AND PRACTICE: A RESOURCE GUIDE 26 (2020), <https://perma.cc/JRQ2-AXHS> (defining “True Positives” and “False Positives”).

140. See HAMILTON, *supra* note 140, at 30 (“[T]he AUC is a discrimination index that represents the probability that a randomly selected recidivist received a higher risk classification than a randomly selected non-recidivist.”).

141. See *id.* (“AUCs range from 0 to 1.0, with .5 indicating no better accuracy than chance, and a 1.0, meaning perfect discrimination (i.e., all recidivists were classified higher than all non-recidivists).”).

142. See Zhang et al., *supra* note 31, at 173, 181, 184; COMPAS Guide, *supra* note 4, at 20-21.

143. See, e.g., CASEY ET AL., *supra* note 2, at 16 (stating that “RNA instruments classify offenders into groups of low, moderate, and high risk of recidivism”); COMPAS Guide, *supra* note 4, at 8 (showing three classification levels for each COMPAS scale type); MULTI-HEALTH SYS., *supra* note 55, at 2 (showing four LSI-R classification levels).



different score ranges can have different true positive and negative rates.<sup>144</sup> An instrument could have a 70% overall success rate but be more accurate for people with low to medium risk scores than those with high risk scores. It could also be more accurate for some subgroups than others; for example, an instrument could have better predictive accuracy for men than women or for one ethnic group than others.<sup>145</sup>

AUC can also be misleading because it only measures how well an instrument distinguishes recidivists from non-recidivists (discrimination),<sup>146</sup> not predictive accuracy (calibration).<sup>147</sup> Melissa Hamilton provides a useful hypothetical of a ten-point instrument that achieves perfect accuracy (AUC = 1.0) in development because all recidivists scored a 2 and all non-recidivists scored a 1.<sup>148</sup> The AUC tells us nothing about the instrument's predictive validity for future subjects who score in the 3-10 range. Additionally, AUC does not tell us the likelihood of recidivism for each risk group; there is no way to tell from an AUC value whether the recidivism rate of a "high risk" group is, say, 30% or 80%.<sup>149</sup>

b. Correlation coefficients ( $r$ )

A correlation coefficient ( $r$ ) measures the strength and direction of a relationship between two variables.<sup>150</sup> Studies use different kinds of correlation coefficients depending on the nature of the variables and data,<sup>151</sup> but coefficients commonly used in risk assessment literature always take a

---

144. See *infra* Part II.B.1 (showing that the OST had far better predictive accuracy in the 5-17 score range than outside that range for a large sample of Virginia probationers).

145. See HAMILTON, *supra* note 140, at 31 (explaining that "AUC metrics on entire populations . . . may obscure difference in AUC levels for subgroups").

146. See *id.* at 25 (defining discrimination and calibration).

147. See *id.* at 31 ("Unfortunately, the AUC is too commonly misinterpreted as measuring calibration accuracy; but a higher AUC does not mean more accurate prospective prediction.").

148. *Id.* at 32.

149. See Seena Fazel, Matthias Burghart, Thomas Fanshawe, Sharon Danielle Gil, John Monahan & Rongqin Yu, *The Predictive Performance of Criminal Risk Assessment Tools Used at Sentencing: Systemic Review of Validation Studies*, 81 J. CRIM. JUST. 101902, at 5 (2022), <https://perma.cc/PWH2-66TA> ("High risk may relate to an actual reoffending rate of 30% or 70%—we do not know using an AUC on its own.").

150. See CASEY ET AL., *supra* note 2, at 16.

151. See Kelly M. Babchishin & Leslie-Maaik Helmus, *The Influence of Base Rates on Correlations: An Evaluation of Proposed Alternative Effect Sizes with Real-World Data*, 48 BEHAV. RES. 1021, 1022 (2015) (explaining that researchers determine the appropriate correlation coefficient based on factors including the nature of the variables and sample size).

value between -1 and 1.<sup>152</sup> An  $r$  of -1 indicates a perfect inverse relationship, an  $r$  of 0 indicates no relationship, and an  $r$  of 1 indicates a perfect positive relationship.<sup>153</sup>

Interpreting  $r$  values is tricky. They measure the strength of a correlation, but that word “strength” is more slippery than it seems. One issue is that an  $r$  value has no objective meaning.<sup>154</sup> A second, related issue is that the threshold for a strong effect size is subjective.<sup>155</sup> One study may consider  $r = 0.38$  to be a “Large” correlation while another considers it a “Weak” correlation.<sup>156</sup> And the interpretive issues do not stop there.

$r$  values do not necessarily have better predictive utility as they move from 0 toward 1 or -1. To illustrate that point, imagine you are investigating the relationship between arm freckles and academic performance. You administer a seven-point quiz and find that for every five freckles, scores rise one point. Because the correlation between the two variables is 100% and the slope is positive,  $r = 1$ .<sup>157</sup>

---

152. See CASEY ET AL., *supra* note 2, at 16 (describing the point-biserial correlation coefficient); Patrick Schober, Christa Boer & Lothar A. Schwarte, *Correlation Coefficients: Appropriate Use and Interpretation*, 126 ANESTHESIA & ANALGESIA 1763, 1766 (2018) (describing the Spearman coefficient); DAVID FARABEE, SHELDON ZHANG, ROBERT E.L. ROBERTS & JOY YANG, SEMEL INST. FOR NEUROSCIENCE & HUM. BEHAV., COMPAS VALIDATION STUDY: FINAL REPORT 23-24 (2010) (using the Spearman coefficient in a COMPAS risk score–recidivism analysis).

153. See Haldun Akoglu, *User's Guide to Correlation Coefficients*, 18 TURK. J. EMERGENCY MED. 91, 91 (2018) (“Zero means there is no correlation, where 1 means a complete or perfect correlation. The sign of the  $r$  shows the direction of the correlation.”).

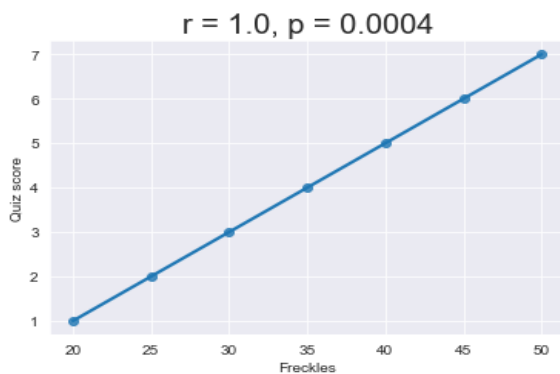
154. Schober et al., *supra* note 153, at 1765 (noting that “[t]he correlation coefficient is sometimes criticized as having no obvious intrinsic interpretation”).

155. See *id.* (“Several approaches have been suggested to translate the correlation coefficient into descriptors like ‘weak,’ ‘moderate,’ or ‘strong’ relationship . . . . These cutoff points are arbitrary and inconsistent and should be used judiciously.”).

156. See CASEY ET AL., *supra* note 2, at 17 (showing an interpretive convention that an  $r$  value exceeding .371 is “Large”); Schober et al., *supra* note 153, at 1765 (showing a “[c]onventional” interpretive approach in which correlations between 0.10 and 0.39 are “Weak”).

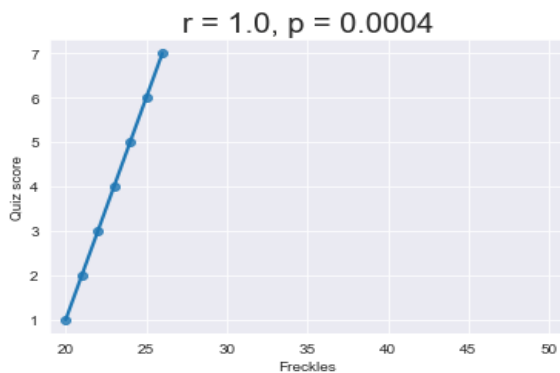
157. The following freckle examples employ a correlation coefficient called Kendall’s tau. See Babchishin & Helmus, *supra* note 152, at 1022 (explaining that Kendall’s tau is appropriate for two ordinal (discrete and ordered) variables and a small sample size).

**Figure 9A**  
Freckle vs. Quiz Score Illustration ( $r = 1.0$ , broad domain)



Now imagine you test a second group of subjects with a lower average number of freckles but still find a perfect correlation. For this second group, quiz scores rise one point for every single additional freckle.

**Figure 9B**  
Freckle vs. Quiz Score Illustration ( $r = 1.0$ , narrow domain)

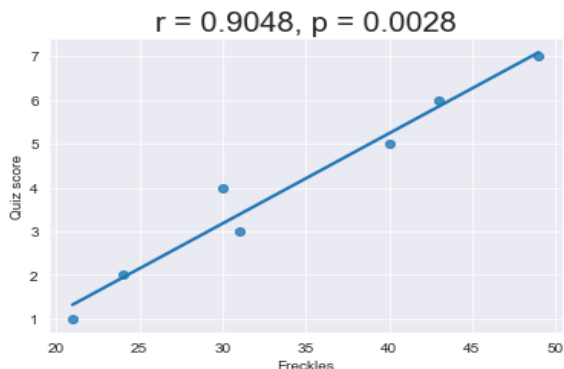


$r = 1$  here too, but the model has less predictive utility than the first. It only gives score predictions for people with fewer than twenty-six freckles on their arms. So you decide to assemble a third test group, aiming for greater freckle diversity. This time, you do not get a perfect correlation, but your model,

nonetheless, has greater predictive utility than when you tested it on the group with a smaller average number of freckles.

**Figure 9C**

Freckle vs. Quiz Score Illustration ( $r = 0.9$ , broad domain)



The freckle examples show how a correlation can be “strong” without being especially useful.  $r$  values may not tell us how well a model based on one sample will perform on a sample with different characteristics. Equally important, as demonstrated above,  $r$  values do not indicate the magnitude of change one variable invokes in the other, but only how uniformly they track together.<sup>158</sup>

There are many other caveats to the reliability and utility of an  $r$  value. One caveat arises when a variable is dichotomous, meaning it can take on only two values.<sup>159</sup> Correlations with a dichotomous variable are “base rate sensitive.”<sup>160</sup> Base rates indicate how often a behavior or characteristic (like recidivism or freckles) occurs in a sample.<sup>161</sup> When a correlation is base rate sensitive, its value changes depending on the average occurrence of that

158. See Schober et al., *supra* note 153, at 1764 fig.1 (demonstrating that for graphs of continuous variables with a linear relationship, “the scatter approaches a straight line as the coefficient approaches -1 or +1”).

159. See Kelly M. Babchishin & Leslie-Maaike Helmus, *The Influence of Base Rates on Correlations: An Evaluation of Proposed Alternative Effect Sizes with Real-World Data*, 48 BEHAV. RSCH. METHODS 1021, 1021 (2015) (describing a “yes/no” item as a “dichotomous” variable).

160. *Id.* at 1022.

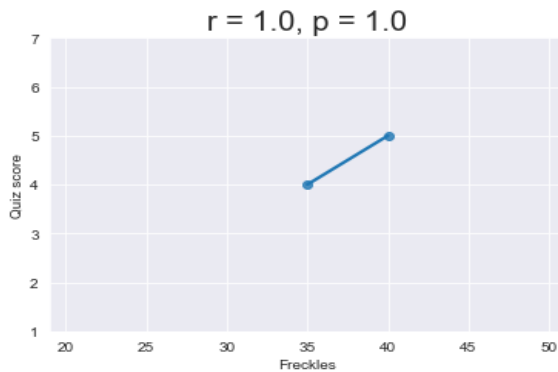
161. BONTA & ANDREWS, *supra* note 47, at 28 (“A base rate is the typical or expected occurrence of behavior.”).

behavior or characteristic.<sup>162</sup> The closer a group's average recidivism rate gets to the extremes (0% or 100%), the harder it is to use risk scores to predict an individual's recidivism risk.<sup>163</sup>

Another caveat is that  $r$  values need to be interpreted in light of statistical significance (the probability that a calculated correlation is a fluke).<sup>164</sup> A sample with only two data points will yield an  $r$  of 1.0 (perfect correlation), for example, but the model has no statistical significance at all ( $p = 1.0$ )<sup>165</sup> and, therefore, no predictive utility. On the flip side, a correlation could have excellent statistical significance but only a nominal  $r$  value.<sup>166</sup>

**Figure 9D**

Freckle vs. Quiz Score Illustration ( $r = 1.0$ , two data points)



- 
162. See Babchishin & Helmus, *supra* note 160, at 1022 (“[T]here are complications with correlations computed for dichotomous data. . . . The issue is that the size of the correlation is influenced by the distribution of the dichotomous variable (i.e., the base rate).”).
163. *Id.* (“[T]he correlation’s sensitivity to base rates has some interpretive meaning, reflecting the reduced statistical power and reliability of the correlation when base rates deviate from 50%. . . . If dichotomous predictions are being made (e.g., this offender is ‘dangerous’ or ‘not dangerous’), then the likelihood of false positives increases with low base rate events. This will be reflected in reduced correlations.” (citations omitted)).
164. See JAMES H. STOCK & MARK W. WATSON, *INTRODUCTION TO ECONOMETRICS* 71 (3d ed. 2017) (“The p-value, also called the significance probability, is the probability of drawing a statistic at least as adverse to the null hypothesis as the one you actually computed in your sample, assuming the null hypothesis is correct.”).
165. See Akoglu, *supra* note 154, at 92 (“The p-value shows the probability that [strength of correlation] may occur by chance.”).
166. See *id.* (“A statistically significant correlation does not necessarily mean that the strength of the correlation is strong.”); Schober et al., *supra* note 153, at 1765 (“With large datasets, very small correlation coefficients can be ‘statistically significant.’ Therefore, a statistically significant correlation must not be confused with a clinically relevant correlation.”).

There are yet further caveats to  $r$  values,<sup>167</sup> but suffice to say that  $r$ , like AUC must be interpreted holistically, not in isolation. Sample sizes, base rates, levels of statistical significance, population subgroup characteristics, and researchers' sophistication vary. Both AUC and  $r$  values should be taken with a grain of salt.

B. No One Has a Good Handle on How to Assess Criminogenic Needs

1. AUC and  $r$  tell an incomplete story

Understanding the relationship between risk factors and recidivism requires ongoing data collection and sophisticated, holistic analysis; it is not just a matter of calculating AUC or  $r$  values, or pulling figures from a meta-analysis.<sup>168</sup> Take Virginia's 2016 OST revalidation results. The Virginia OST revalidation report includes charts with the number of probationers and average recidivism rates for each OST score, which facilitates data visualization and analysis.<sup>169</sup> In the graph below, dot size represents the number of Virginia probationers who received each OST score, ranging from five probationers (score of 30) to 571 probationers (score of 9).<sup>170</sup> None of the 6,852 probationers in the sample scored higher than 30 on the OST.<sup>171</sup>

---

167. See, e.g., Schober et al., *supra* note 153, at 1765 (describing assumptions about linearity, monotonicity, distribution, and independence that must be met to appropriately employ particular correlation coefficients).

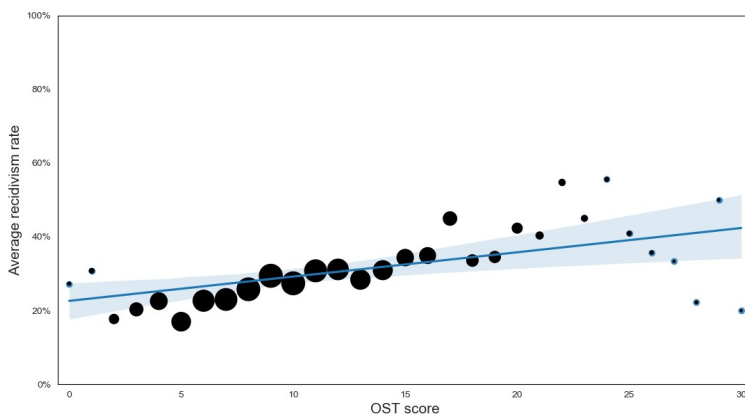
168. Parts II.B.-C owe a debt to Christopher Baird's National Council on Crime and Delinquency special report critiquing conventional approaches to criminal risk assessment. See generally CHRISTOPHER BAIRD, NAT'L COUNCIL ON CRIME & DELINQ., A QUESTION OF EVIDENCE: A CRITIQUE OF RISK ASSESSMENT MODELS USED IN THE JUSTICE SYSTEM 9 (2009), <https://perma.cc/X66A-RD72> (drawing on numerous studies of risk assessment instruments to challenge common practices in instrument design and measurement of predictive validity, reliability, and effectiveness).

169. NAT'L CTR. FOR STATE CTS., *supra* note 80, at 9-10, app. B at 26.

170. *Id.* at 9-10.

171. *Id.*

**Figure 10**  
Virginia OST Revalidation Results



A couple of observations:

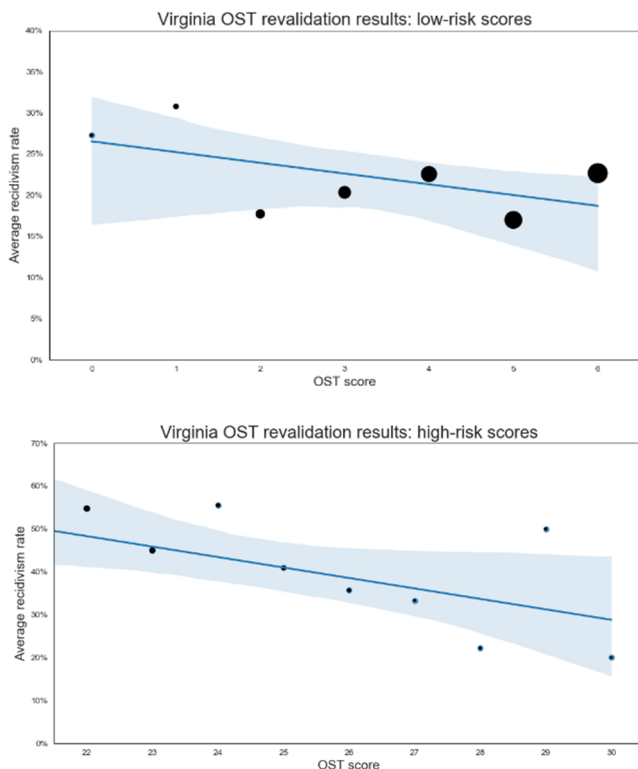
First, the OST was reasonably good at sorting probationers in the sample into three risk groups, but the groups' base rates only ranged from 21% to 44%. The low risk group (score  $\leq 6$ ) had a 21% average recidivism rate, the medium risk group (score of 7-20) had a 30% recidivism rate, and the high risk group (score  $\geq 21$ ) had a 44% recidivism rate.<sup>172</sup> For practical purposes like sentencing, there is not much to differentiate an individual with a 21% risk from an individual with a 30% risk, or an individual with a 30% risk from one with a 44% risk. Moreover, it is unintuitive that 44% should be a threshold for "high risk," since more than half of the high risk group did *not* recidivate.

Second, according to these results, the OST cannot produce meaningful predictions or need assessments outside the 5 to 17 score range. The 18% ( $n = 1,243$ )<sup>173</sup> of probationers who fell outside that range did not show a strong, positive correlation between score and recidivism; in fact, recidivism trended *downward* as scores rose for both outer groups.

172. See *id.* at 19, 26 (providing OST score, recidivism, and risk group data).

173. See *id.* at 9-10 (providing OST score, recidivism, and risk group data).

**Figure 11**  
Virginia OST Revalidation Results: Low- and High-Risk Scores



The way the instrument falls apart at the outer edges may be because too few subjects scored outside the 5-17 score range to demonstrate a representative trend, but it intimates that factors have different degrees of influence at different risk levels. The OST factors, as weighted by the instrument, bear a positive correlation to the behavior of defendants who score in the 5-17 range, but not those who score higher or lower. For that reason, if the OST can be said to identify Virginia probationers' "needs" at all, it may only do so within the 5-17 score range.

The Virginia results support the common agreement among scholars and policymakers that risk assessment instruments must be normed and periodically revalidated on the population of each jurisdiction that uses



them.<sup>174</sup> Revalidation can turn up unforeseen instrument limitations, like how the OST had good predictive validity only for medium-scoring probationers in the Virginia sample. Revalidation is also necessary because different populations, whether distinguished by geography, gender, race, offense type, or other factors, have different average crime rates and risk-need profiles.<sup>175</sup> An instrument calibrated to Virginia probationers, for example, might end up looking quite different than an instrument calibrated to misdemeanor defendants in another state. Predictive validity is fluid and complex; it cannot be established by slapping an *r* or AUC value on an instrument after one or two validation studies.

## 2. The Gendreau meta-analysis is not an oracle

Leading scholars and risk-assessment developers frequently cite a 1996 meta-analysis, enthusiastically titled “A Meta-Analysis of the Predictors of Adult Offender Recidivism: What Works!,” as a source of authority for the relative strengths of recidivism risk factors.<sup>176</sup> The meta-analysis has substantial caveats and limitations, which are acknowledged by its authors but not the scholars who continue to cite it.<sup>177</sup> The LSI-R’s developers present it as

---

174. See COLLINS ET AL., *supra* note 71, at 6 (“[V]alidation of the instrument on the local offender population is considered the best practice, particularly when extensive research on that instrument has not been conducted.”); CASEY ET AL., *supra* note 2, at 22 (explaining the kinds of changes that commonly occur over time and necessitate revalidation).

175. See *infra* note 337 (describing different need scale results for different races and ethnicities); FIELD OPERATIONS ADMIN., *supra* note 31, at 10 (explaining that COMPAS has gender-specific scales because studies have shown women to have a different need profile than men); Duwe, *supra* note 136, at 584 (“[R]ecent research suggests that disaggregating offender criminal histories by the type and timing of the offense may be helpful in further increasing predictive accuracy.”); Brandon L. Garrett & John Monahan, *Judging Risk*, 108 CALIF. L. REV. 439, 490 (2020) (“Instruments should be revalidated over time at reasonable intervals and with attention to local variation in populations, resources, and crime patterns.”).

176. Paul Gendreau, Tracy Little & Claire Goggin, *A Meta-Analysis of the Predictors of Adult Offender Recidivism: What Works!*, 34 CRIMINOLOGY 575 (1996). The title is a response to an influential 1974 article by Robert Martinson that surveyed the existing literature on offender rehabilitation and argued there was no sound evidence that any approach worked. See Robert Martinson, *What Works?—Questions and Answers About Prison Reform*, PUB. INT., Spring 1974, at 21, 25 (“With few and isolated exceptions, the rehabilitative efforts that have been reported so far have had no appreciable effect on recidivism. Studies that have been done since our survey was completed do not present any major grounds for altering that original conclusion.” (emphasis omitted)). For examples of scholarship citing to the Gendreau meta-analysis, see, for example, *infra* note 182.

177. See, e.g., Gendreau et al., *supra* note 177, at 587 (acknowledging that the authors “did not attempt to adjust statistically for methodological artifacts, which may or may not have had an impact on the magnitude of the effect sizes obtained” and that the groups surveyed were “regrettably” homogenous in terms of gender, race, and nationality);  
*footnote continued on next page*

a significant source of empirical support for their “central eight” theory of criminal behavior,<sup>178</sup> the COMPAS Field Guide cites to the Gendreau study as gospel twelve times,<sup>179</sup> and ORAS’s developers cite it as a source of authority as well.<sup>180</sup> Much of the scholarship cited in this Article also takes for granted that the study successfully determined which factors have a consistent and substantial bearing on recidivism.<sup>181</sup>

Gendreau and his co-authors evaluated one hundred and thirty-one recidivism studies in an attempt to identify the strongest predictors of recidivism.<sup>182</sup> They found that several dynamic risk factors showed at least as strong a correlation with recidivism as any static factor, outside of criminal history.<sup>183</sup>

**Table 4**<sup>184</sup>

Gendreau Meta-Analysis Mean Effect Sizes for Predictor Domains

| Static Factors                           | <i>r</i> |
|--|----------|
| Age                                      | .15      |
| Criminal History: Adult                  | .18      |
| History of Antisocial Behavior: Preadult | .13      |
| Family Criminality                       | .12      |
| Family Rearing Practices                 | .15      |
| Family Structure                         | .10      |
| Gender                                   | .10      |
| Intellectual Functioning                 | .07      |
| Race                                     | .13      |
| SES                                      | .06      |

*table continued on next page*

---

BONTA & ANDREWS, *supra* note 47, at 53-54 (providing an uncritical discussion of findings from the meta-analysis).

178. BONTA & ANDREWS, *supra* note 47, at 53-54.

179. COMPAS Guide, *supra* note 4, at 20, 26-27, 32, 36, 37, 39, 42, 45, 48 & 49.

180. LATESSA ET AL., *supra* note 61, at 7.

181. For propositions citing to the meta-analysis in these articles, see, for example, Duwe, *supra* note 136, at 580 (“Interventions should also target the known dynamic predictors of recidivism . . . .”); and Hamilton, *supra* note 7, at 237 (“Dynamic factors incorporate criminogenic needs, which are often mutable in nature, and, therefore, may be proper targets for rehabilitative programming.”).

182. Gendreau et al., *supra* note 177, at 581 (“We identified 131 studies as suitable for a meta-analysis.”); *id.* at 579 (explaining that the meta-analysis addresses the questions of “[w]hich predictor domains predict recidivism, and are some more potent than others?”).

183. *See id.* at 583.

184. *Id.*

| Dynamic Factors        | r   |
|------------------------|-----|
| Antisocial Personality | .18 |
| Companions             | .18 |
| Criminogenic Needs     | .18 |
| Interpersonal Conflict | .15 |
| Personal Distress      | .05 |
| Social Achievement     | .15 |
| Substance Abuse        | .14 |

The authors believed they had identified “which predictor domains and actuarial measures of risk will be most useful to practitioners and policymakers.”<sup>185</sup> They expressed confidence that “[a]dditional research, in our view, is not likely to change the direction or ordering of the results of the predictor domains to any marked degree”<sup>186</sup>—a bold statement that does not hold up under scrutiny.

The meta-analysis is an interesting, intellectually honest study, but it is no oracle and should not be given the benefit of the doubt. As an initial matter, scholars have found evidence of authorship bias in instrument validation studies, including those for the LSI-R.<sup>187</sup> Equally important, the meta-analysis does not interrogate whether the included instruments use similar methodology to assess antisocial personality, companions, and “criminogenic needs,” which it deems three of the four most significant predictors of recidivism.<sup>188</sup> If a meta-analysis lumps together instruments that define a factor differently, the findings cannot necessarily be imputed back out to all instruments. Parts III.B.3 and IV.D of this Article discuss evidence that

---

185. *Id.* at 591.

186. *Id.* at 588.

187. Studies authored by instrument developers tend to find better predictive validity than studies published by independent researchers. See Jay P. Singh, Martin Grann & Seena Fazel, *Authorship Bias in Violence Risk Assessment? A Systematic Review and Meta-Analysis*, 8 PLOS ONE e72484, at 6 (2013), <https://perma.cc/42BC-DF2F> (“Evidence of a significant authorship effect was found, specifically to risk assessment studies published in peer-reviewed journals.”); Zhang et al., *supra* note 31, at 184-85; Holly A. Wilson & Leticia Gutierrez, *Does One Size Fit All?: A Meta-Analysis Examining the Predictive Ability of the Level of Service Inventory (LSI) with Aboriginal Offenders*, 41 CRIM. JUST. & BEHAV. 196, 206 (2014) (“Significantly greater predictive validity was found when the studies were conducted by a LSI developer.”); Michael S. Caudy, Joseph M. Durso & Faye S. Taxman, *How Well Do Dynamic Needs Predict Recidivism? Implications for Risk Assessment and Risk Reduction*, 41 J. CRIM. JUST. 458, 460 (2013) (“Most of the empirical support for the big four and central eight as direct predictors of recidivism outcomes has come from the creators of the RNR model and their colleagues.”).

188. Gendreau et al., *supra* note 177, at 582-83.

instruments that purport to assess the same factor actually assess different things, indicating that Gendreau’s ranking of factors does not hold universally.

The meta-analysis’s treatment of “criminogenic needs” should also lessen confidence in the findings. The authors define “criminogenic needs” as “antisocial attitudes supportive of an antisocial life-style and behavior regarding education, employment.”<sup>189</sup> That definition is difficult to parse and could cover a wide range of question types. It is also out of step with most scholarship, which typically defines “criminogenic needs” as synonymous with dynamic risk factors.<sup>190</sup> The confusion caused by the authors’ idiosyncratic definition of “criminogenic needs” is compounded by their choice to differentiate it from “Antisocial Personality,” “Interpersonal Conflict,” and “Personal Distress.”<sup>191</sup> Those are not commonly agreed-upon categories. There is no intuitive way for the LSI-R emotional and cognitive categories, for example, to map onto the Gendreau categories.

**Table 5**  
LSI-R and Gendreau Categories Comparison

| LSI-R<br>Emotional/Personal<br>Category <sup>192</sup>  | LSI-R<br>Attitudes/Orientation<br>Category <sup>193</sup>   | Gendreau et al. Cognitive<br>and Emotional Predictor<br>Domains <sup>194</sup>  |
|---|---|---|
| <ul style="list-style-type: none"> <li>• Moderate interference</li> <li>• Severe interference, active psychosis</li> <li>• Mental health treatment–past</li> <li>• Mental health treatment–present</li> <li>• Psychological assessment indicated</li> </ul> | <ul style="list-style-type: none"> <li>• Supportive of crime</li> <li>• Unfavorable toward convention</li> <li>• Poor toward sentence</li> <li>• Poor toward supervision</li> </ul> | <ul style="list-style-type: none"> <li>• Antisocial personality</li> <li>• Criminogenic needs</li> <li>• Interpersonal conflict</li> <li>• Personal distress</li> </ul> |

---

189. *Id.* at 597.

190. *See, e.g.,* BONTA & ANDREWS, *supra* note 47, at 13 (“The predictors [of criminal conduct] are called risk factors, and when those risk factors are dynamic (subject to change), they are called dynamic risk factors (or criminogenic needs).”); JAMES, *supra* note 128, at 3 (“Dynamic risk factors, also called ‘criminogenic needs,’ change and/or can be addressed through interventions.”).

191. Gendreau et al., *supra* note 177, at 583.

192. Idaho LSI-R Scoring Guide—Version 3.0, *supra* note 57, at 23-25.

193. *Id.* at 26-27.

194. Gendreau et al., *supra* note 177, at 583.

The meta-analysis authors must have made many compromises and judgment calls to assemble a common factor classification system across instruments.<sup>195</sup>

Recidivism metrics are also a weak point in the meta-analysis. Recidivism can be measured in several different ways.<sup>196</sup> The meta-analysis includes studies that variously used “arrest, conviction, incarceration, parole violation, or a combination thereof” as outcome variables.<sup>197</sup> Different risk assessment instruments do not necessarily measure the same thing if their recidivism metrics are different. The relationship between a risk factor and rearrest, for example, may be different than the relationship between that same factor and parole violations. The authors acknowledge as much, noting a significant difference among the mean effect sizes for different outcome variables, with mean *r* values for incarceration “significantly greater than those of conviction or parole violation.”<sup>198</sup> Which recidivism metric a study uses, in other words, can be outcome-determinative.

The Gendreau meta-analysis covered a set of populations that (a) can be expected to have a broad range of recidivism base rates and (b) are non-representative of a high percentage of criminal defendants. The various study samples include Danish adoptee property offenders, male psychopaths, youth heroin offenders, Kalamazoo probation enhancement program participants, hyperactive boys and their brothers, and patients returned to hospitals for the criminally insane.<sup>199</sup> The authors described the set of studies they analyzed as “regrettably, virtually silent” on the prediction of recidivism among nonwhite and non-male subjects:<sup>200</sup>

The data base was, regrettably, virtually silent on the prediction of recidivism among female offenders, minority groups, white-collar offenders, and some important sample characteristics, such as risk level and the psychological make-up of the subjects studied. Much of the effect size data on dynamic predictor domains came from Canada, where there has been a strong emphasis on the assessment of individual differences.<sup>201</sup>

Not only were the study samples nonrepresentative of many populations, but each domain was analyzed across a different number and combination of

---

195. A further shortcoming of the Gendreau meta-analysis is that it does not list all of the instruments employed by the covered studies.

196. See CASEY ET AL., *supra* note 2, at 9 (listing various ways instrument developers have defined recidivism).

197. Gendreau et al., *supra* note 177, at 579.

198. *Id.* at 586 n.2.

199. *Id.* at 598-607.

200. *Id.* at 587.

201. *Id.*

studies. The number of effect sizes<sup>202</sup> per domain ranged from 17 (gender) to 282 (criminal history—adult plus pre-adult).<sup>203</sup>

Meta-analyses find commonalities across many groups, but that does not mean a meta-analysis’s results can be imputed to every group—all the more so when the authors disclose that its dataset is “virtually silent” on everyone but white male defendants.<sup>204</sup> The most the meta-analysis can tell us is that a correlation exists between certain factors and recidivism for some populations. The exact effect sizes mean very little.

The compromises Gendreau and his co-authors made are justified in a meta-analysis but should cast doubt on the robustness and generalizability of their conclusions. COMPAS’s reliance on the Gendreau findings is particularly confounding, as equivalent’s Practitioner’s Guide admits that “there is no agreement on the particular attitudinal dimensions or cognitions that are the most useful for predictive purposes” and grants that “[c]learly, this area could require a highly extensive inventory to map the full range of cognitive dimensions relative to crime.”<sup>205</sup> Leading scholars are even less critical than COMPAS’s developers. One article on constitutional issues in evidence-based sentencing cites the meta-analysis 49 times in 76 pages.<sup>206</sup> What evidence the article cites to support the Gendreau findings is largely in the vein of other scholars agreeing that the factors are criminogenic, not that the Gendreau team measured them well.<sup>207</sup> The meta-analysis should not be cited that way. It does not offer definitive guidance on how to quantify or rank risk factors.

### 3. Dynamic factors are only weakly correlated with crime

Risk assessment factors are only useful if they are substantially correlated with recidivism, irrespective of whether they are classified as “risk” or “need” factors. The purpose of need scores is to craft a sentence, program assignment, or set of release terms that reduce the risk a person will recidivate.<sup>208</sup> If a factor

---

202. Some studies reported more than one effect size per predictor domain. *See id.* at 581, 583 (reporting that 131 studies were identified as suitable for the meta-analysis and including predictor domains where the number of effect sizes exceeded 131).

203. *Id.* at 583-84.

204. *Id.* at 587.

205. COMPAS Guide, *supra* note 4, at 39-40.

206. Oleson, *supra* note 20, at 1337, 1347, 1350-54, 1356, 1359-67, 1381, 1384, 1388, 1394.

207. *See, e.g., id.* at 1353-54 (supporting Gendreau’s finding on “criminal companions”); *id.* at 1354 (stating that “[n]umerous studies have related antisocial attitudes with criminality”); *id.* at 1359 (citing to “a considerable body of work” that agrees with Gendreau’s finding on pre-adult “antisocial behavior”).

208. *See, e.g., ELEK ET AL., supra* note 25, at 4 (“The need principle maintains that treatment services should target an offender’s dynamic risk factors or criminogenic needs to reduce an offender’s probability of recidivism.”); *Malenchik v. State*, 928 N.E.2d 564, *footnote continued on next page*

is not strongly correlated with recidivism, then interventions targeting that factor are unlikely to be very effective.

The Gendreau results would not reveal a strong or causal relationship between any factor and recidivism even if they were generalizable to all populations. Recall that  $r$  measures how well a factor tracks with recidivism. According to a common convention in risk assessment scholarship, for a sample with a 50% base rate of recidivism, the  $r$  cutoffs for small, medium, and large effect sizes are 0.100, 0.243, and 0.371.<sup>209</sup> That convention is generous; in fields such as medicine, politics, and psychology, researchers often interpret any  $r$  value below 0.3 as weak or poor.<sup>210</sup> Even accepting the more generous convention, though, studies consistently find that most static (inalterable) factors and all dynamic (changeable) factors are only weakly correlated with recidivism.

No factor in the Gendreau meta-analysis exceeded  $r = 0.18$  except for a weighted estimate of “Companions” ( $r = 0.21$ ).<sup>211</sup> The meta-analysis’s methodology makes it possible to convert  $r$  to AUC.<sup>212</sup> At a 50% base rate, an  $r$  of 0.18 equals an AUC of approximately 0.60 and an  $r$  of 0.22 equals an AUC of approximately 0.63, both of which are quite low (recall that an AUC of 0.60 means there is a 60% chance that a randomly chosen recidivist has a higher score on the scale than a randomly chosen non-recidivist).<sup>213</sup> The meta-analysis does not show that any factor has a particularly strong correlation with recidivism, much less that any correlation is causal.

---

573 (Ind. 2010) (“[E]vidence-based assessment instruments can be significant sources of valuable information for judicial consideration in deciding whether to suspend all or part of a sentence, how to design a probation program for the offender, whether to assign an offender to alternative treatment facilities or programs, and other such corollary sentencing matters.”).

209. See Marnie E. Rice & Grant T. Harris, *Comparing Effect Sizes in Follow-Up Studies: ROC Area, Cohen’s  $d$ , and  $r$* , 29 LAW & HUM. BEHAV. 615, 617 (2005); CASEY ET AL., *supra* note 2, at 17. Those cutoffs, which correspond to Cohen’s  $d$  values of 0.2, 0.5, and 0.8, do not change much unless the base rates get pretty extreme. Using the Rice-Harris conversion formula for a sample with a 29% recidivism base rate, like the Virginia OST revalidation sample, those same cutoffs are 0.09, 0.22, and 0.34.

210. See Schober et al., *supra* note 153, at 1765; Akoglu, *supra* note 154, at 92.

211. Gendreau et al., *supra* note 177, at 583 tbl.1. No weighted estimates other than “companions” exceeded  $r = 0.18$  except for the weighted composite “Risk Scales” measure, which was the same as the unweighted composite measure at  $r = 0.30$ . *Id.*

212. Rice and Harris provide a conversion formula and chart for the Pearson product-moment correlation coefficient (Pearson’s  $r$ ), Cohen’s  $d$ , and AUC, when one of the variables is dichotomous. See Rice & Harris, *supra* note 210, at 616, 619. The Gendreau meta-analysis estimated effect sizes using Pearson’s  $r$ , with recidivism coded as a dichotomous variable. *Id.* at 579-80.

213. See *id.* at 616, 619; COMPAS Guide, *supra* note 4, at 13 (asserting that “[t]he consensus in the field of recidivism research seems to be that AUC values below 0.65 are poor”).

Many studies beyond the Gendreau meta-analysis have found only weak effect sizes for dynamic factors. A 2013 study of 24,972 LSI-R subjects led by criminologist Michael Caudy separated out each factor in the big four (antisocial personality, antisocial cognition, antisocial associates, and criminal history) and found only a modest correlation between the dynamic factors and recidivism.<sup>214</sup>

**Table 6**<sup>215</sup>

Caudy Study—Dynamic Factors

|                      | Sample 1 |     | Sample 2 |     |
|----------------------|----------|-----|----------|-----|
|                      | <i>r</i> | AUC | <i>r</i> | AUC |
| Antisocial attitudes | .14      | .58 | .01      | .50 |
| Antisocial peers     | .12      | .56 | .11      | .56 |
| Family/marital       | .07      | .54 | .03      | .52 |

Criminal history, however, bore a far stronger correlation with recidivism than the dynamic factors in the big four and a modestly stronger correlation than the total LSI-R risk score.

**Table 7**<sup>216</sup>

Caudy Study—Criminal History and Total LSI-R Score

|                   | Sample 1 |     | Sample 2 |     |
|-------------------|----------|-----|----------|-----|
|                   | <i>r</i> | AUC | <i>r</i> | AUC |
| Criminal history  | .24      | .65 | .17      | .60 |
| Total LSI-R score | .23      | .63 | .15      | .59 |

These results cast substantial doubt on the contribution of dynamic factors to LSI-R predictive accuracy. The low *r* values of the dynamic factors also suggest that even if the attitudes and relationships assessed by the LSI-R have some causal relationship with recidivism, the causality is limited, at best.

Consistent with the Gendreau and Caudy studies, two COMPAS validation studies found that no factor on its own has a particularly strong correlation with recidivism. The studies reported *r* and AUC values for the General

214. Caudy et al., *supra* note 188, at 463. All results, including criminal history and total LSI-R score, were statistically significant at  $p < .0001$  except for antisocial attitudes and family/marital in Sample 2, which were not statistically significant even at  $p < .05$ . *Id.* at 464 tbl.6.

215. *Id.* at 462 tbl.4.

216. *Id.*



Recidivism Risk Scale and need scales.<sup>217</sup> One study found that no need scale had an AUC greater than 0.63.<sup>218</sup> The other study found an AUC of 0.66 for one composite need scale (Criminal Opportunity) but AUCs of 0.64 or less for the other seventeen need scales it evaluated.<sup>219</sup> The COMPAS Guide asserts a recidivism research consensus that “AUC values below 0.65 are poor, 0.65 to 0.69 are fair, 0.70 to 0.75 are good, and 0.76 and above are excellent.”<sup>220</sup> By that token, every need scale but one performed poorly, and the exception was a composite scale that did not even achieve a “good” AUC value.

For COMPAS, the low *r* and AUC values of individual factors are a problem hiding in plain sight. *equivant* does not claim that need scales are strong predictors of risk.<sup>221</sup> COMPAS need scale results, however, are often labeled “criminogenic” in score reports.<sup>222</sup>

*equivant* should not set a lower bar for need factors than for risk factors. The COMPAS Practitioner’s Guide asserts that an AUC of 0.60 for its Criminal Associates Scale would be “a modest result if this were a standalone risk scale,” but “the result indicates good criterion validity” for a need scale.<sup>223</sup> The distinction between risk and need factors is specious, though. Most instruments, including those discussed in this Article, incorporate both into risk score calculations and/or present need factors as criminogenic. When factors are only weakly correlated with recidivism, they will neither improve predictive validity nor identify appropriate targets for intervention. It does not matter whether factors are labeled “risk,” “need,” or “dynamic.” Weak correlation makes a factor unhelpful for both risk and needs assessment.<sup>224</sup>

#### 4. Call for item-level transparency

Risk assessment literature tends to focus on categories, not the items within each category, which makes it difficult to identify precisely which attitudes and

---

217. COMPAS Guide, *supra* note 4, at 23-24. One study examined all nineteen need scales and the other examined eighteen need scales but omitted the composite Cognitive Behavioral scale. *Id.* at 23-24, 36.

218. *Id.* at 23.

219. *Id.* at 24.

220. *Id.* at 13.

221. *See supra* note 36.

222. *See, e.g.,* Fordham Parole Info. Project, *supra* note 41, at 2 (titling the need scale report “Criminogenic Need Scales”); COMPAS Questionnaire, *supra* note 32 (titling the need scale report “Criminogenic and Needs Profile”).

223. COMPAS Guide, *supra* note 4, at 22.

224. *See* BAIRD, *supra* note 169, at 3 (“If a factor on a risk assessment instrument has no demonstrated relationship to recidivism, how can positive changes in this factor reduce risk?”).

associations are correlated with recidivism.<sup>225</sup> Individual items might have no predictive value or, worse, *lessen* the predictive validity of their section. The best way to test whether an item adds value is to compare the section’s correlation with recidivism when the item is present and when it is removed.

Suppose you are validating a “Criminal Attitudes” instrument section, consisting of two items, on a sample of two hundred subjects. The first item asks whether the subject loves committing crimes and the second asks whether the subject loves vanilla ice cream. You, the researcher, observe that one hundred subjects answered “yes” to “Do you love committing crimes?” and of those hundred, ninety-nine recidivated. One hundred subjects also answered “yes” to “Do you love vanilla ice cream?” but of those hundred, only fifty recidivated.

**Table 8**  
Criminal Attitudes

|                 | Do you love committing crimes? | Do you love vanilla ice cream? |
|-----------------|--------------------------------|--------------------------------|
| Recidivists     | 99 yes, 1 no                   | 50 yes, 50 no                  |
| Non-recidivists | 1 yes, 99 no                   | 50 yes, 50 no                  |

The ice cream item has no predictive value, but the “Criminal Attitudes” section as a whole does. The likelihood that any given recidivist answers “yes” to either question is 99.5%.<sup>226</sup> The likelihood any given non-recidivist answers “yes” to either question is 50.5%.<sup>227</sup> Et voilà, the section as a whole is great at identifying subjects whose criminal attitudes place them at higher risk of recidivism. You assign equal weight to both questions because that makes the math easy. Now, a future subject who loves vanilla ice cream will receive a higher risk score than one who does not, even though the question has no predictive value. If you had excluded the vanilla ice cream item, though, you would have a section that can predict with nearly 100% accuracy whether a subject will recidivate.

This example may seem outlandish, but the OST, LSI-R and ORAS-CST also assign equal weight to items that are unlikely to be equally correlated with recidivism.<sup>228</sup> The OST and LSI-R, for instance, assign the same point value to

225. See, e.g., David J. Simourd & Jennifer Van De Ven, *Assessment of Criminal Attitudes: Criterion-Related Validity of the Criminal Sentiments Scale-Modified and Pride in Delinquency Scale*, 26 CRIM. JUST. & BEHAV. 90, 100 tbl.2 (1999) (reporting effect sizes for subscales but not individual items).

226.  $\frac{99}{100} + \frac{50}{100} - \left(\frac{99}{100} * \frac{50}{100}\right) = 99.5\%$ .

227.  $\frac{1}{100} + \frac{50}{100} - \left(\frac{1}{100} * \frac{50}{100}\right) = 50.5\%$ .

228. See STINSON, *supra* note 81, at 98-102; LSI-R Scoresheet, *supra* note 3; UNIV. OF CINCINNATI CTR. FOR CRIM. JUST. RSCH., *supra* note 62, at 2-9.

attitudes “supportive of crime” and attitudes “unfavorable to convention” or “non-conforming to societal norms” in their “Attitude” sections.<sup>229</sup> There is no reason to take it on faith that each of those items is equally correlated with recidivism or that each truly contributes to the predictive validity of the section as a whole.

When advocates call for transparency, they should call for it at the item level. Scholars and policymakers have overlooked many problems with attitude and associational factors because they tend to focus on whether particular categories of questions are morally or legally defensible. The more granular questions about whether the items within each category are morally, legally, or empirically defensible are equally important.

### C. More Factors, More Problems

There are two rationales for quantifying attitudes and associations: (a) because those factors improve an instrument’s accuracy, and (b) because they provide information that is useful to judges, probation officers, correctional officers, and parole boards.<sup>230</sup> Wrong on both counts. They do neither.

#### 1. Attitude and associational factors do not improve predictive accuracy

Instruments like the LSI-R, OST, and ORAS-CST that include attitude and associational factors in risk scores have not been shown to outperform those that exclude them. In fact, many models have achieved AUCs or predictive accuracy equal to that of leading instruments with as few as three factors.

- A 2018 study found that both a simple algorithm and a cohort of layperson Mechanical Turks<sup>231</sup> achieved an AUC of 0.71 with only seven static

---

229. STINSON, *supra* note 81, at 101 (OST “Attitude” items); Idaho LSI-R Scoring Guide—Version 3.0, *supra* note 57, at 26-27 (LSI-R “Attitudes/Orientation” items).

230. See Megan R. Kopkin, Stanley L. Brodsky & David DeMatteo, *Risk Assessment in Sentencing Decisions: A Remedy to Mass Incarceration?*, 9 J. AGGRESSION, CONFLICT & PEACE RSCH. 155, 158 (2017) (“If an assessment is only to be used to characterize an individual’s behaviour, an absence of dynamic factors is not particularly problematic. However, if the intention of the risk assessment is to manage or reduce risk, the inclusion of dynamic variables is integral.”).

231. The original Mechanical Turk was a fraudulent chess-playing automaton with a human hidden inside. See generally TOM STANDAGE, *THE TURK: THE LIFE AND TIMES OF THE FAMOUS EIGHTEENTH-CENTURY CHESS-PLAYING MACHINE* (2002) (recounting the history of the Mechanical Turk). The authors of the 2018 study recruited scorers from Amazon’s Mechanical Turk, “an online crowdsourcing marketplace where people are paid to perform a wide variety of tasks.” Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, 4 SCI. ADVANCES eao5580, at 4 (2018), <https://perma.cc/LP5B-MUUY>.

factors.<sup>232</sup> That same study claims to have modestly bested COMPAS in overall predictive accuracy by running a logistic regression on only two factors: age and total number of prior convictions.<sup>233</sup>

- A 2014 study of approximately 25,000 California parolees found that COMPAS's General Recidivism scale achieved an AUC of 0.70,<sup>234</sup> while a scale with only four factors—gender, age, age at first arrest, and number of prior arrests—achieved an AUC of 0.72.<sup>235</sup>
- A 2018 study employing a machine learning model matched COMPAS's accuracy with only gender, age, and number of prior crimes as inputs.<sup>236</sup>
- In the 2016 OST and MOST revalidation study on 16,799 Virginia community-based probationers, National Center for State Courts (NCSC) researchers compared the predictive performance of the Virginia MOST and a proxy risk score.<sup>237</sup> The Virginia MOST had eight static and dynamic items, while the proxy risk score had three static items.<sup>238</sup> The researchers recommended replacing the MOST with a proxy score because the latter was quicker to complete and had “better utility and increased accuracy.”<sup>239</sup>
- A 2008 effort by leading criminologists to develop a new pretrial risk assessment instrument tested sixty-three items across eight domains.<sup>240</sup> The researchers found that only six factors across three domains (criminal history, drug use, and employment) were significantly associated with failure to appear and rearrest while on pretrial release.<sup>241</sup>
- The Public Safety Assessment (“PSA”), a non-proprietary instrument adopted by many states for pretrial risk assessment, utilizes nine factors:

---

232. Dressel & Farid, *supra* note 232, at 2.

233. *Id.* at 4 (displaying an overall accuracy rate of 66.8% for the authors' two-feature classifier, contrasted with an overall accuracy rate of 65.4% for COMPAS). The authors do not state which COMPAS scale they used for comparison. Criminologists pointed out in a rejoinder to the study that “although COMPAS does collect information on a large number of dynamic risk factors, relatively few of these contribute to the instrument's overall risk score.” Holsinger et al., *supra* note 5, at 53. That argument, however, only supports the exclusion of need factors from risk assessment.

234. The COMPAS General Recidivism scale has at least five inputs. *See* COMPAS Guide, *supra* note 4, at 31.

235. *See* Zhang et al., *supra* note 31, at 181-84.

236. Elaine Angelino, Nicholas Larus-Stone, Daniel Alabi, Margo Seltzer & Cynthia Rudin, *Learning Certifiably Optimal Rule Lists for Categorical Data*, 18 J. MACH. LEARNING RSCH. art. 234, at 40-42, 68-69 (2018), <https://perma.cc/AWL6-TM2P>.

237. NAT'L CTR. FOR STATE CTS., *supra* note 80, at 4.

238. *Id.* at 12, 20-21.

239. *Id.* at 17.

240. Christopher T. Lowenkamp, Richard Lemke & Edward Latessa, *The Development and Validation of a Pretrial Screening Tool*, FED. PROB., Dec. 2008, at 2, 4.

241. *Id.* at 4-5.

age at current arrest, six criminal history factors, and two failure-to-appear factors.<sup>242</sup> The PSA's developers evaluated hundreds of factors and chose those nine because they had the greatest predictive abilities.<sup>243</sup>

Research from the past twenty years, including the Caudy study discussed above in Part II.B.3, supports the proposition that the LSI-R would benefit from PSA-style pruning. A 2003 LSI-R validation and reliability study commissioned by the Pennsylvania Commission on Crime and Delinquency found that “[o]nly eight of the 54 LSI-R items were . . . associated with recidivism.”<sup>244</sup> Those items were related to criminal and arrest history, residential instability, drug problems, and mental health treatment.<sup>245</sup> The authors found that if they pared the LSI-R down to those eight factors, the instrument's accuracy improved.<sup>246</sup> The excised factors, they explained, introduced noise into the model both because they were uncorrelated with recidivism and because they had relatively high levels of disagreement among scorers.<sup>247</sup> A 2004 analysis of the Y-LSI (a close cousin to the LSI-R) similarly found that “relatively few of the 42 items contribute to accuracy in risk classification.”<sup>248</sup> A 2012 Pennsylvania Commission on Sentencing study analyzed the risk factors used by the LSI-R and several other instruments and selected only “nine significant predictors of recidivism,” which lay within the domains of age, gender, criminal history, county of residence, and current offense type.<sup>249</sup>

It is unsurprising that predictive models with dozens of factors do not outperform those with only a few, given the perils of overfitting and the inevitable multicollinearity of criminogenic factors. “Overfitting” occurs when researchers take a kitchen-sink approach to statistical analysis by including too many factors and/or testing a model on too small a dataset.<sup>250</sup> The resultant

---

242. Arnold Ventures, Public Safety Assessment FAQs (“PSA 101”) 2 (2019), <https://perma.cc/G8XE-Z6SG>.

243. Stanford L. Sch. Pol’y Lab, Risk Assessment Factsheet: Public Safety Assessment (PSA) 2 (2019), <https://perma.cc/LZ6A-HYV2>.

244. AUSTINET AL., *supra* note 58, at iii.

245. *Id.* at 18-20.

246. *See* BAIRD, *supra* note 169, at 4-5.

247. AUSTINET AL., *supra* note 58, at 18, 22.

248. Anthony W. Flores, Lawrence F. Travis III & Edward J. Latessa, Case Classification for Juvenile Corrections: An Assessment of the Youth Level of Service/Case Management Inventory (YLS/CMI), Executive Summary 1 (2004), <https://perma.cc/C2PU-KBY6>.

249. Kopkin et al., *supra* note 231, at 157.

250. *See* Jake Lever, Martin Krzywinski & Naomi Altman, *Model Selection and Overfitting*, 13 NATURE METHODS 703, 703 (2016) (explaining that overfitting is a “common pitfall” in building and evaluating a predictive regression model, “whereby the model not only fits the underlying relationship between variables in the system . . . but also fits the

*footnote continued on next page*

model adopts the idiosyncrasies of the original dataset, such that it does not generalize well to other datasets.<sup>251</sup> “Multicollinearity” means that when you build a statistical model with a bunch of highly correlated factors—such as poverty, neighborhood crime, and family criminal history—the model loses the ability to precisely estimate the relative contribution of each factor to an outcome like recidivism.<sup>252</sup> Consequently, piling on factors can reduce a model’s predictive accuracy.

The studies cited above strongly counsel parsimony in instrument construction, yet developers, both public and private, keep piling on the factors. Many, in fact, believe that strong correlations between factors prove an instrument’s reliability.<sup>253</sup> That makes little sense. If both smoking and taking ibuprofen increase heart attack risk, it does not follow that a heart attack risk assessment is more reliable if there is a strong correlation between smoking and taking ibuprofen. Strong correlations among factors are not a virtue, but a red flag for overfitting and multicollinearity.<sup>254</sup>

## 2. “Need” factors should be explained, not quantified

It is entirely appropriate for decisionmakers to *consider* need factors, but it is inappropriate to quantify them. Even an instrument like COMPAS, which (probably) includes attitude and associational factors only in “need” scores, still implies that “criminogenic needs” are readily quantifiable and have a

---

noise unique to each observed sample”); Xue Ying, *An Overview of Overfitting and Its Solutions*, 1168 J. PHYSICS: CONF. SERIES art. 022022, at 1, (2019), <https://perma.cc/2RF2-WG6W> (explaining that overfitting can be caused by too small a training set or too many inputs).

251. See Ying, *supra* note 251, at 1; see also Tracy L. Fass, Kirk Heilbrun, David DeMatteo & Ralph Fretz, *The LSI-R and the COMPAS: Validation Data on Two Risk-Needs Tools*, 35 CRIM. JUST. & BEHAV. 1095, 1098 (2008) (noting that COMPAS may have achieved an AUC of 0.79 in its initial validation study because it was developed and validated on the same small dataset, and that the results may not be generalizable because the study did not analyze race or gender).
252. STOCK & WATSON, *supra* note 165, at 199-203 (explaining that imperfect multicollinearity occurs when “two or more . . . regressors are highly correlated” and “implies that it will be difficult to estimate precisely one or more of the partial effects using the data at hand”).
253. See, e.g., COMPAS Guide, *supra* note 4, at 28 (“Generally, if the items entering a scale are highly correlated (internally consistent), then the summated scale will be reliable.”); STINSON, *supra* note 81, at 41 (“Items were included in the OST based on a belief that all of them are factors that contribute to an individual’s risk to re-offend. If this is true, all of the items in the OST should be related to each other.”).
254. Cf. BAIRD, *supra* note 169, at 8 (“For [criminal] risk assessment, it is best when all risk items are totally independent of each other but each has a relatively strong relationship to the outcome measure utilized.”).

demonstrated, substantial correlation with the commission of crime.<sup>255</sup> Neither implication is correct.

The information that underlies need scores should already be in pretrial, presentence, and pre-parole reports, in narrative form; quantifying that information obscures more than it illuminates.<sup>256</sup> The obscurity problem can be observed, for example, in a 2012 OST report, in which the defendant scored a 5 (out of 8 possible points) in the “Family and Social Relationships” domain, but the narrative explained only that he was separated from his wife and reportedly received divorce papers while in jail.<sup>257</sup> The report rendered the impression that the defendant had significant social problems without supporting evidence. Consider also the words of a Wisconsin judge confronted by several counterintuitive COMPAS results:

A second area where the . . . information simply doesn’t correspond with the COMPAS is in your cognitive behavior. In the COMPAS it says that you have an absence of blaming others, making excuses, or minimizing the seriousness of the offense. All you have done throughout this entire process is to blame [the victim] and to make excuses for what happened.<sup>258</sup>

The COMPAS score was not just unhelpful; it directly contradicted something that was obvious in the record.

Quantification implies a strong, statistically significant relationship between a factor and recidivism, even where none exists, and obscures the subject’s individual characteristics and experiences.<sup>259</sup> Qualitative information removes a layer of obscurity between the subject’s responses and the report the judge receives. Simply put, a risk assessment instrument that quantifies a defendant’s “needs” does not improve on pure narrative and pure human discretion.

---

255. *See supra* Part I.A.

256. The Indiana Court of Appeals said as much in 2008:

Here, the trial court used the LSI-R score as an aggravator in addition to performing an independent evaluation of the evidence. This is also problematic, because areas analyzed in this psychological inventory appear duplicative of factors already considered by the trial court in sentencing (criminal history, education, employment) and other areas appear of questionable value (leisure and recreation).

*Rhodes v. State*, 896 N.E.2d 1193, 1195 (Ind. Ct. App. 2008).

257. *Presentence Investigation at 4, State v. Motari*, No. CR2010 0-005514-001-DT (Ariz. Super. Ct. Feb. 27, 2012).

258. Brief of Defendant-Appellant Frederick A. Ziesmer at 15, *State v. Ziesmer*, No. 2016AP1889 (Wis. Ct. App. Dec. 7, 2016) (quoting the trial court).

259. In 1983, Justice Blackmun expressed this concern in a dissent to a Supreme Court decision that deemed psychiatric expert testimony predicting future dangerousness admissible even if it is wrong “most of the time.” *Barefoot v. Estelle*, 463 U.S. 880, 901 (1983). His dissent noted that the use of psychiatric categories, which have “little or no demonstrated relationship to violence[,] . . . often obscures the unimpressive statistical or intuitive bases for prediction.” *Id.* at 932 (Blackmun, J., dissenting).

Categorization is useful even though quantification is not. In a 2002 evaluation of evidence-based sentencing in Virginia, researchers found that judges and probation officers tended to like how risk assessment instruments offered a consistent framework for consideration of relevant factors.<sup>260</sup> Like a good legal standard, the framework kept judges focused on the most useful and similar factors in every case; however, judges sentenced “in accord with risk assessment recommendation[s]” only in about two-thirds of cases.<sup>261</sup> The judges and probation officers found the standardization useful, not the quantification. Court systems should keep the categories but jettison the quantification of need factors.

### III. Speech and Expressive Association

#### A. Abstract Beliefs are Inadmissible at Sentencing Under the First Amendment

##### 1. *Dawson* requires a nexus

It has been a matter of settled law since the Supreme Court decided *Dawson v. Delaware* in 1992 that a defendant’s abstract beliefs and expressive associations may not be considered at sentencing unless they have some bearing on the issue being tried.<sup>262</sup>

David Dawson burglarized a home and killed its occupant following a prison escape in 1986.<sup>263</sup> At the penalty phase of Dawson’s capital proceedings, the judge permitted the prosecution to introduce evidence showing that Dawson belonged to the Aryan Brotherhood.<sup>264</sup> The only information about the gang stipulated to the jury was: “The Aryan Brotherhood refers to a white racist prison gang that began in the 1960’s in California in response to other gangs of racial minorities. Separate gangs calling themselves the Aryan

---

260. BRIAN J. OSTROM, MATTHEW KLEIMAN, FRED CHEESMAN II, RANDALL M. HANSEN & NEAL B. KAUDER, NAT’L CTR. FOR STATE CTS., OFFENDER RISK ASSESSMENT IN VIRGINIA: A THREE-STAGE EVALUATION 2 (2002), <https://perma.cc/WPS8-VUHP>.

261. *Id.* at 3.

262. 503 U.S. 159, 168 (1992) (“[The First Amendment] prevents Delaware here from employing evidence of a defendant’s abstract beliefs at a sentencing hearing when those beliefs have no bearing on the issue being tried.”). The standard established in *Dawson* applies equally to speech and expressive association, as the opinion invokes the two rights in interchangeable fashion. *Id.* at 167 (“Dawson’s First Amendment rights were violated by the admission of the [associational] evidence in this case, because the evidence proved nothing more than Dawson’s abstract beliefs.”).

263. *Id.* at 160-61.

264. *Id.* at 162.



Brotherhood now exist in many state prisons including Delaware.”<sup>265</sup> The jury recommended the death sentence.<sup>266</sup>

The Supreme Court reversed Dawson’s death sentence, finding that the admission of the Aryan Brotherhood evidence violated his First Amendment rights because it “proved nothing more than Dawson’s abstract beliefs.”<sup>267</sup> Chief Justice Rehnquist, writing for the majority, elaborated that since both Dawson and his victim were white and the jury did not hear evidence about the Aryan Brotherhood’s promotion of violent escape and murder, the prosecution did not properly link the association and the crime.<sup>268</sup> Chief Justice Rehnquist explained that the state “might have avoided this problem if it had presented evidence showing more than mere abstract beliefs on Dawson’s part, but on the present record one is left with the feeling that the Aryan Brotherhood evidence was employed simply because the jury would find these beliefs morally reprehensible.”<sup>269</sup>

The Court rejected the government’s argument that under Delaware law, Dawson’s abstract beliefs were admissible to demonstrate his “character.”<sup>270</sup> The Delaware sentencing statute at issue permitted a jury to consider “all relevant evidence in aggravation or mitigation” relating to the “character and propensities” of the defendant at sentencing.<sup>271</sup> The Court regarded the government’s argument as a matter of semantics, not substance. “Whatever label is given to the evidence presented,” the Court noted, “we conclude that Dawson’s First Amendment rights were violated by the admission of the Aryan Brotherhood evidence in this case, because the evidence proved nothing more than Dawson’s abstract beliefs.”<sup>272</sup>

*Dawson* imposed a constitutional limit on the use of character evidence at sentencing that overrides the legal precedents and federal and state sentencing statutes granting courts carte blanche. Prior Supreme Court rulings took a more expansive stance on the use of character evidence at sentencing, expressing that “a judge may appropriately conduct an inquiry broad in scope, largely unlimited either as to the kind of information he may consider, or the source from which it may come”<sup>273</sup> and that “the possession of the fullest information possible concerning the defendant’s life and characteristics” is

---

265. *Id.*

266. *Id.* at 163.

267. *Id.* at 167.

268. *See id.* at 165-66.

269. *Id.* at 167.

270. *Id.* at 166-67.

271. *Id.* at 170 (Thomas, J., dissenting) (citing DEL. CODE ANN. tit. 11, § 4209(d)(1) (1987)).

272. *Id.* at 167 (majority opinion).

273. *United States v. Tucker*, 404 U.S. 443, 446 (1972).

“[h]ighly relevant—if not essential.”<sup>274</sup> Some federal and state sentencing statutes are similarly expansive.<sup>275</sup> *Dawson*, however, imposed a limit on the “largely unlimited”<sup>276</sup> sentencing inquiry consistent with other constitutional protections that cabin a court’s discretion, including the proscription of sentences based on a prior successful appeal of conviction or “materially untrue” assumptions about a defendant’s criminal record.<sup>277</sup> Constitutional constraints apply, no matter the breadth of statutory language.<sup>278</sup>

The limit established by *Dawson* is not absolute, though. *Dawson* asserted that the Constitution “does not erect a *per se* barrier to the admission of evidence concerning one’s beliefs and associations at sentencing simply because those beliefs and associations are protected by the First Amendment.”<sup>279</sup> The opinion granted that associational evidence will, in many cases, “serve a legitimate purpose in showing that a defendant represents a future danger to society.”<sup>280</sup> The *Dawson* Court did not, however, delimit the extent to which evidence of beliefs or associations can go to future dangerousness, offering, by way of explanation, only an extreme example that “[a] defendant’s membership in an organization that endorses the killing of any identifiable group . . . might be relevant to a jury’s inquiry into whether the defendant will be dangerous in the future.”<sup>281</sup>

Several circuits have held that associations and beliefs are admissible to prove future dangerousness when there is a properly established nexus

---

274. *Williams v. New York*, 337 U.S. 241, 247 (1949). Although *Tucker* and *Williams* were decided before the Federal Sentencing Guidelines were enacted, the Supreme Court has continued to cite them as sources of authority for the breadth of the sentencing inquiry. See, e.g., *Pepper v. United States*, 562 U.S. 476, 487-89 (2011); *Dawson*, 503 U.S. at 164.

275. See 18 U.S.C. § 3661 (“No limitation shall be placed on the information concerning the background, character, and conduct of a person convicted of an offense which a court of the United States may receive and consider for the purpose of imposing an appropriate sentence.”); *Dawson*, 503 U.S. at 170 (Thomas, J., dissenting). Justice Thomas’s dissent highlights the import of *Dawson*’s departure from prior, more permissive doctrine. “Until today,” Justice Thomas wrote, “we have never hinted that the First Amendment limits the aspects of a defendant’s character that [courts and juries] may consider.” *Id.* at 177.

276. *Tucker*, 404 U.S. at 446.

277. *North Carolina v. Pearce*, 395 U.S. 711, 723-24 (1969) (prohibiting vindictive resentencing); *Townsend v. Burke*, 334 U.S. 736, 740-41 (1948) (prohibiting sentencing based on “materially untrue” assumptions).

278. See *Dawson*, 503 U.S. at 167-68.

279. *Id.* at 165.

280. *Id.* at 166.

281. *Id.*

between the character evidence and the crime.<sup>282</sup> The Aryan Brotherhood affiliation in *Dawson* was not, after all, inadmissible per se, but inadmissible as used by the prosecution. The Court explained that it may have been admissible to prove an aggravating circumstance if the prosecution proved “that the Aryan Brotherhood had committed any unlawful or violent acts, or had even endorsed such acts.”<sup>283</sup> In 1997, the Fifth Circuit took up that very issue in *Fuller v. Johnson* and held that evidence of an Aryan Brotherhood affiliation did not violate the First Amendment when the government, to demonstrate future dangerousness, used it to show that a convicted murderer “was a member of a gang that had committed unlawful or violent acts.”<sup>284</sup> The Fifth Circuit elaborated that the government “did not merely stipulate that Fuller was in the Aryan Brotherhood,” but introduced evidence of the gang’s acts of violence that could lead a reasonable juror to “conclude that membership in such a gang is relevant to future dangerousness.”<sup>285</sup> The Aryan Brotherhood affiliation was not admissible in *Dawson* to give the jury some generalized sense of bad character, but it was admissible in *Fuller* to convince the jury that the defendant was likely to commit another violent crime because he belonged to a violent gang.<sup>286</sup>

The nexus is the key ingredient in *Dawson*-admissible evidence, not the strength of an attitude or association’s correlation with recidivism. If mere relevance to future dangerousness was the standard, then any speech or associational evidence correlated with recidivism would be admissible, and *Dawson* would have no teeth. If the prosecution could produce a study demonstrating that people who think “Teletubbies” is scintillating television are 90% more likely to commit crimes, a defendant’s love of Tinky Winky and friends would, nonetheless, be inadmissible at sentencing under *Dawson*. If the prosecution could prove Trump supporters were more likely to commit crimes than the general population, a defendant’s political leanings would still be inadmissible at sentencing. If electronic dance music (EDM) fans are statistically more likely to commit crimes than the general population—which is a reasonable assumption, solely on the basis that the fan base skews young<sup>287</sup>—an EDM-loving defendant’s musical tastes would also be inadmissible at sentencing.

---

282. See, e.g., *Toomer v. Att’y Gen. U.S.*, 810 Fed. App’x 147, 151 (3d Cir. 2020); *United States v. Simkanin*, 420 F.3d 397, 417-18 (5th Cir. 2005); *Kapadia v. Tally*, 229 F.3d 641, 647-48 (7th Cir. 2000).

283. *Dawson*, 503 U.S. at 166.

284. 114 F.3d 491, 498 (5th Cir. 1997).

285. *Id.*

286. See *id.*

287. See Oleson, *supra* note 20, at 1361-62 (detailing a common consensus among criminologists that age is inversely correlated with the commission of crime).

And so on. *Dawson* requires more than mere correlation: It requires a nexus between evidence of future dangerousness and offense.

2. Factors in general risk assessment instruments lack a nexus to the defendant's offense

*Dawson* should pose a formidable barrier to the use of attitude factors at sentencing that are not tailored to the specific defendant and crime. General risk assessment instruments ask the same questions of all defendants, irrespective of the crime at issue.<sup>288</sup> They crystallize defendants' abstract opinions into a risk score without establishing the nexus required under *Dawson* between their responses and the crime.

COMPAS's "attitude" items violate the *Dawson* standard that "a defendant's abstract beliefs, however obnoxious to most people, may not be taken into consideration by a sentencing judge."<sup>289</sup> The COMPAS questionnaire includes a "Criminal Attitudes" section in which the subject must check off "Strongly Disagree," "Disagree," "Not Sure," "Agree," or "Strongly Agree" in response to eleven statements, including "When people get into trouble with the law it's because they have no chance to get a decent job," "The law doesn't help average people," and "Many people get into trouble or use drugs because society has given them no education, jobs or future."<sup>290</sup> Those statements express abstract beliefs about social issues.

Assessments that rely, or relied, on the 1999 OST questionnaire also violate *Dawson*. The OST questionnaire instructs the subject to answer "Yes" or "No" to statements and questions including "I think it is okay to have tattoos or bodypiercing," "I believe that society is to blame for many problems in the world," and "Sometimes a person has to do whatever it takes to make money, even if is not exactly what would be considered 'normal.'"<sup>291</sup> Those are all abstract beliefs insufficiently particularized to any the crime to pass muster under *Dawson*. When a predictive instrument includes such questions, its use at sentencing is patently unconstitutional.

3. "It's Just One Factor" does not cure First Amendment harms

A First Amendment challenge to evidence-based sentencing could succeed where a Fourteenth Amendment challenge would not because the Fourteenth Amendment allows for attenuation in the mix of factors, while the First does not. Attenuation was the crux of the two most comprehensive judicial

---

288. See *supra* Part I.

289. *Wisconsin v. Mitchell*, 508 U.S. 476, 485 (1993).

290. COMPAS Questionnaire, *supra* note 32, at 8.

291. STINSON, *supra* note 81, at 93-94.

opinions on evidence-based sentencing to date, handed down by the Wisconsin Supreme Court in 2016 and Indiana Supreme Court in 2010.<sup>292</sup> The Wisconsin decision stated that “[a]lthough it cannot be determinative, a sentencing court may use a COMPAS risk assessment as a relevant factor” for decisions on matters such as diversion to alternative sanctions, community supervision, and terms of probation.<sup>293</sup> The Indiana decision similarly stated that LSI-R and Substance Abuse Subtle Screening Inventory (SASSI) scores “may, and if possible should, be considered to supplement and enhance a judge’s evaluation, weighing, and application of the other sentencing evidence in the formulation of an individualized sentencing program appropriate for each defendant.”<sup>294</sup> This argument, dubbed “It’s Just One Factor” by Sonja Starr,<sup>295</sup> has spurred several commentators to compare the use of racial proxies in risk assessment instruments to the use of race in university admissions.<sup>296</sup> “It’s Just One Factor” can cure a constitutional harm in a Fourteenth Amendment context—hence, the university admissions comparison—but there can be no argument under *Dawson* that it cures a First Amendment harm.

The *Dawson* Court rejected the state’s argument that it was “entitled to introduce any ‘bad’ character evidence” because the defense had introduced “‘good’ character evidence.”<sup>297</sup> This argument “misse[d] the mark” because the Aryan Brotherhood evidence could not be “viewed as relevant ‘bad’ character evidence in its own right.”<sup>298</sup> *Dawson*-prohibited speech and association cannot be used at sentencing at all, regardless of what other factors are in play.

Folding abstract beliefs into a broader risk-needs inquiry does not cure but *perpetuates* First Amendment harms, because judges typically do not see the underlying questions when they are presented with risk scores in presentence reports.<sup>299</sup> An Arizona judge, for example, would not know if a defendant received a point in the “Attitude” domain because he thinks it is “okay to commit crime” or because he thinks it is “okay to have tattoos or bodypiercing.”<sup>300</sup>

---

292. See *State v. Loomis*, 881 N.W.2d 749, 768 (Wis. 2016); *Malenchik v. State*, 928 N.E.2d 564, 575 (Ind. 2010).

293. *Loomis*, 881 N.W.2d at 767.

294. *Malenchik*, 928 N.E.2d at 573.

295. Sonja B. Starr, *The New Profiling: Why Punishing Based on Poverty and Identity Is Unconstitutional and Wrong*, 27 FED. SENT’G REP. 229, 233 (2015).

296. See *Grutter v. Bollinger*, 539 U.S. 306, 334 (2003) (holding that universities may “consider race or ethnicity . . . flexibly as a ‘plus’ factor in the context of individualized consideration of each and every applicant”); Sidhu, *supra* note 16, at 697-98; Oleson, *supra* note 19, at 1384-88; Hamilton, *supra* note 7, at 258-59.

297. *Dawson v. Delaware*, 503 U.S. 159, 167-68 (1992).

298. *Id.* at 168.

299. See *supra* note 26 and accompanying text.

300. See STINSON, *supra* note 81, at 93.

**Figure 12**<sup>301</sup>  
2015 OST “Attitude” Assessment

---

---

| DOMAIN   | SCORE |     | SUMMARY  |
|----------|-------|-----|--|
| Attitude | 1/7   | 14% | <ul style="list-style-type: none"><li>• Attitudes at odds with societal norms.</li><li>• Amenable to probation.</li><li>• Respectful attitude toward authority figures.</li><li>• Appears motivated to improve life.</li></ul> |

---

---

As Starr has written, “if a court’s decisionmaking is unconstitutional in substance, it surely cannot become constitutional through obscurity of reasoning.”<sup>302</sup> Judges unwittingly violate the Constitution when they sentence based on obscured unconstitutional factors.

Attitude risk assessment items should be subject to a “Would You Put it on the Record” test, in accordance with Starr’s admonition that “obscurity of reasoning” does not cure constitutional harms.<sup>303</sup> If a judge stated on the record that her sentence was informed by the defendant’s agreement with propositions like “I think it is okay to have tattoos or bodypiercing” or “many people get into trouble or use drugs because society has given them no education, jobs, or future,” her decision would invite a *Dawson* challenge. If an item would be inadmissible when extracted from the instrument and put on the record, it does not belong in an instrument used at sentencing.

## B. Normative Issues in Attitude Assessment

### 1. What even is a criminal attitude?

Risk assessment literature variously refers to cognition and opinions correlated with crime as “criminal attitudes,”<sup>304</sup> “procriminal attitudes,”<sup>305</sup> “criminal thinking,”<sup>306</sup> “attitudes supportive of an antisocial lifestyle,”<sup>307</sup> and “antisocial attitudes.”<sup>308</sup> The range of terminology signals that many types of

---

301. Presentence Investigation at 3, *State v. Rall*, No. CR2015-107409-001-DT (Ariz. Super. Ct. Sept. 22, 2015).

302. Starr, *supra* note 22, at 863-64.

303. *Id.*

304. LATESSA ET AL., *supra* note 61, at 17.

305. Hamilton, *supra* note 7, at 247.

306. COMPAS Guide, *supra* note 4, at 40.

307. Duwe, *supra* note 136, at 580.

308. Hamilton, *supra* note 7, at 235.

attitudes could fit the bill. Bonta and Andrews, developers of the LSI-R and co-authors of the textbook *The Psychology of Criminal Conduct*, describe “procriminal attitudes” as follows:

This set of variables includes attitudes, values, beliefs, rationalizations, and thoughts that are favorable to crime. The cognitive-emotional states associated with crime are anger and feeling irritated, resentful, and/or defiant. Specific indicators would include identification with criminals, negative attitudes toward the law and justice system, a belief that crime will yield rewards, and rationalizations that specify a broad range of conditions under which crime is justified (e.g. the victim deserved it, the victim is worthless).<sup>309</sup>

Some of Bonta and Andrews’ indicators are more obviously conducive to crime than others. Identification with criminals and rationalizations seem more specific to people who have committed crimes than the belief that crime will yield rewards, which is accurate in some instances, or negative attitudes toward the law and justice system, which are quite common in the general population. A May 2021 poll by the National Opinion Research Center, for example, found that 25% percent of all Americans agreed that the criminal justice system “needs a complete overhaul” while another 43% agreed that it “needs major changes.”<sup>310</sup> A 2022 survey commissioned by the NCSC found that 51% of respondents rated state court performance as “fair/poor,” outnumbering the 49% of respondents who rated state court performance as “excellent/good.”<sup>311</sup> A 2019 Pew survey found that 65% of U.S. adults believed that the U.S. criminal justice system “treats black people less fairly.”<sup>312</sup> In surveys commissioned by the NCSC conducted from 2014 to 2017, only about half of respondents agreed that judges make unbiased decisions.<sup>313</sup> Such attitudes could be correlated with crime, but they are also mainstream and not necessarily causal.

The relationship between criminality and some of the opinions in the OST and COMPAS questionnaires is obscure. The OST penalizes acceptance of tattoos and body piercings,<sup>314</sup> which are not only commonplace in the American population, but important traditions in some cultures.<sup>315</sup> The

---

309. BONTA & ANDREWS, *supra* note 47, at 45.

310. Kat Stafford & Hannah Fingerhut, *AP-NORC Poll: Police Violence Remains High Concern in U.S.*, ASSOCIATED PRESS & NORC CTR. FOR PUB. AFFS. RSCH. (May 21, 2021), <https://perma.cc/MPY7-3YTW>.

311. GBAO STRATEGIES, 2022 STATE OF THE STATE COURTS – NATIONAL SURVEY ANALYSIS 2 (2022), <https://perma.cc/AW88-A3RY>.

312. PEW RSCH. CTR., RACE IN AMERICA 2019, at 11 (2019), <https://perma.cc/KJ9J-HWK2>.

313. GBA STRATEGIES, 2017 STATE OF THE COURTS SURVEY ANALYSIS 3 (2017), <https://perma.cc/VG8K-YL5G>.

314. *See* STINSON, *supra* note 81, at 93.

315. *See* PEW RSCH. CTR., HOW YOUNG PEOPLE VIEW THEIR LIVES, FUTURES AND POLITICS: A PORTRAIT OF “GENERATION NEXT” 21 (2007), <https://perma.cc/MWH5-YT5K> (finding a  
*footnote continued on next page*)

COMPAS item “A hungry person has a right to steal” is an opinion as likely to be held by a law-abiding *Les Misérables* fan as by a defendant with a lengthy criminal history. And the OST item “I believe that society is to blame for many problems in the world” is not really an attitude antithetical to convention; the belief that society is to blame for major world problems like war and pollution is a conventional, even if not universal, attitude.

Not only do some risk assessment items seem to bear a shaky relationship with criminality, but some label as “criminal attitudes” what is actually an accurate understanding of the world rooted in many subjects’ personal experiences. A Washington, D.C. public defender challenged a juvenile client’s score on the Structured Assessment of Violence Risk in Youth (SAVRY) instrument in 2017, for instance, and found that he was assigned a higher risk score based, in part, on his “negative attitudes toward the police.”<sup>316</sup> “There are obviously plenty of reasons for a black male teenager to not like police,” the attorney told an *Atlantic* reporter.<sup>317</sup> The OST similarly penalizes people with negative attitudes toward the police.<sup>318</sup> COMPAS penalizes subjects who agree that “[m]any people get into trouble or use drugs because society has given them no education, jobs or future”—a proposition that is not only entirely consistent with many people’s life experiences, but also the foundation of the “strain theory” of criminality.<sup>319</sup> Strain theory posits that people commit crimes because structural inequality blocks their ability to attain the “American dream” through legitimate means.<sup>320</sup> Per the COMPAS Practitioner’s Guide, it is an important criminological theory that “may provide guidance for effective interventions.”<sup>321</sup> It is hardly fair to label antipathy toward the police and subscription to strain theory as “criminal” or “antisocial.”

---

36% tattoo rate and 30% body piercing rate for adults ages 18-25 and a 40% tattoo rate and 22% body piercing rate for adults ages 26-40); Michele Metych, *Tattoo*, ENCYCLOPÆDIA BRITANNICA ONLINE, <https://perma.cc/J9U8-EW5W> (last updated Mar. 17, 2023) (citing ancient and contemporary practices around the world).

316. Derek Thompson, *Should We Be Afraid of AI in the Criminal-Justice System?*, ATLANTIC (June 20, 2019), <https://perma.cc/8JJJ-UZ48>; see Justin Chae, *Hidden Algorithms, Bad Science, and Discrimination in the D.C. Juvenile Justice System*, MEDIUM (July 1, 2020), <https://perma.cc/HT47-GAGC> (identifying the instrument as SAVRY).

317. Thompson, *supra* note 317.

318. STINSON, *supra* note 81, at 94.

319. See COMPAS Guide, *supra* note 4, at 6 (“[Social strain theory] is an economic explanation for crime. Crime occurs largely in poverty-stricken areas where opportunities to attain the ‘American Dream’ by legitimate means is [sic] blocked, producing frustration and a desire to pursue monetary success by any means necessary.”); BONTA & ANDREWS, *supra* note 47, at 38 (describing the evolution of strain theory over time).

320. BONTA & ANDREWS, *supra* note 47, at 38; COMPAS Guide, *supra* note 4, at 6.

321. COMPAS Guide, *supra* note 4, at 5.



## 2. Attitudes as racial proxies

The *Dawson* principle that no one may be punished for their abstract views applies equally to people from all walks of life, but any discussion of viewpoint penalization is incomplete without reference to race. Perhaps the most common criticism of criminal risk assessment is that it amounts, in the words of Bernard Harcourt, to “statistical discrimination.”<sup>322</sup> The close statistical relationship between race and risk factors—such as poverty, arrests, and neighborhood crime—that arises from societal inequities is well-trodden territory, most comprehensively covered in Harcourt’s book, *Against Prediction*.<sup>323</sup> Along with Harcourt, many scholars, journalists, and courts have given careful consideration to the proxy issue, i.e., that the close relationship ensures that even if race is not an explicit factor in risk assessments, it is inevitably an implicit factor.<sup>324</sup> No author, though, has spoken to the proxy issue inherent in “criminal attitude” risk assessment items.

The absence of discussion may be an oversight, but it may also reflect a reasonable reticence to tread into tricky territory. The sole author who has come close to the criminal attitudes proxy issue has been inartful in his otherwise thoughtful article on the limits of fairness as a risk assessment objective.<sup>325</sup> Ben Green describes risk assessment categories such as “Community Disorganization,” “Anger Management Problems,” and “Poor Compliance,” as well as questions like “Is there much crime in your neighborhood?” as features that “resist objective answers, turning these assessments into value-laden affairs in which white, Western, and middle-class standards are imposed on defendants.”<sup>326</sup> His discussion reaches toward something true, which is that many instruments penalize antipathy toward instruments of authority, like law enforcement, the legal system, and civic institutions. Green does not, however, define “white, Western, and middle-class standards” or acknowledge the fact that many criminal defendants are white and middle-class.<sup>327</sup>

---

322. See BERNARD E. HARCOURT, *AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHING IN AN ACTUARIAL AGE* 241 (2007).

323. See generally *id.* (advocating against actuarial methods in law enforcement and criminal sanctions).

324. See, e.g., Starr, *supra* note 22, at 838 (“[T]he socioeconomic and family variables that [instruments] include are highly correlated with race, as is criminal history, so they are likely to have a racially disparate impact.”).

325. See generally Ben Green, *The False Promise of Risk Assessments: Epistemic Reform and the Limits of Fairness*, 2020 PROC. OF THE CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY 594, 601-02, <https://perma.cc/PD93-62EJ> (arguing that risk assessment stakeholders must expand their focus beyond “fairness” to promote systemic reform).

326. *Id.* at 596.

327. See E. ANN CARSON, BUREAU OF JUST. STAT., NCJ NO. 302776, PRISONERS IN 2020—STATISTICAL TABLES 10 (2021), <https://perma.cc/5XMT-XJBF> (reporting that in 2020, 30% of people incarcerated in the United States were white); Press Release, Prison Pol’y  
*footnote continued on next page*

It is dangerously reductive to assume that any person's value system is determined by their race, ethnicity, or financial status, or that particular values or behaviors are exclusive to certain "cultures." The danger becomes clear when one drills down on the specific questions that subjects must answer. The COMPAS questionnaire asks, for example, the extent to which a subject agrees with the statements "[s]ome people just don't deserve any respect and should be treated like animals" and "[s]ome people must be treated roughly or beaten up just to send them a clear message."<sup>328</sup> It is deeply offensive (and, of course, incorrect) to presume that the extent of agreement or disagreement with that statement is a matter of "cultural values." This is not to say that Green makes that presumption, but that framing the discussion in terms of cultural standards poses dual, related risks of reinforcing damaging stereotypes and conflating individual values with group identity.

An LSI-R training manual demonstrates how easily reference to "cultural values" can spill over into outright racism. The manual intimates that respect for the attitudes it labels "antisocial thinking" is a manifestation of naive cultural relativism.<sup>329</sup> In its list of "7 Less Promising Targets for Change," the manual includes "[s]howing respect for antisocial thinking on the grounds that the values of one culture are as equally valid as the values of another culture (when a culture holds criminal values that harm others)."<sup>330</sup> That statement is not explicitly racist, but a brief article included in the manual is. The article asserts as to "Ethnicity/Culture":

This is an area so extensive that I cannot possibly begin to address all the ramifications in this brief paper. However, with minority populations being over represented [sic] in offender populations, it is an element that must be considered. It is my suggestion that you talk to minority co-workers, as a starting point, to get some input.<sup>331</sup>

This paragraph speaks volumes about the author's biases and assumptions: that minority populations may be overrepresented due to their ethnicity or culture rather than societal inequities, that the reader is white, and that it is

---

Initiative, *Prisons of Poverty: Uncovering the Pre-incarceration Incomes of the Imprisoned* (July 9, 2015), <https://perma.cc/6ZME-DZ4A> (reporting that in 2004, 22% of incarcerated men and 15% of incarcerated women had pre-incarceration incomes of at least \$37,500).

328. COMPAS Questionnaire, *supra* note 33, at 8.

329. LSI-R: LEVEL OF SERVICE INVENTORY-REVISED: INTRODUCTORY TRAINING COURSE—PARTICIPANT MANUAL, *supra* note 46, at 15.

330. *Id.*

331. Peter Philbrick, *Responsivity: The Key to Understanding & Maximizing Offender Change*, in LSI-R: LEVEL OF SERVICE INVENTORY-REVISED: INTRODUCTORY TRAINING COURSE—PARTICIPANT MANUAL, *supra* note 46, at 41, 47.

appropriate for white assessors to question their nonwhite co-workers on whether their ethnicity or culture is conducive to criminality.

A more nuanced take on the viewpoint-as-proxy issue is that, assuming subjects answer truthfully, the answers to “criminal attitudes” questions are likely to correlate with the extent of marginalization and disadvantage subgroups have experienced, even though individual responses will not neatly stratify along racial or “cultural” lines. “Criminal attitude” assessments include questions designed to gauge a subject’s attitudes toward other individuals and views on social convention, authority, and institutional legitimacy.<sup>332</sup> “Some people just don’t deserve any respect and should be treated like animals,” for example, gauges attitudes toward others, as does “If someone insults my friends, family or group they are asking for trouble.”<sup>333</sup> “The law doesn’t help average people,” on the other hand, gauges respect for social institutions.<sup>334</sup> Both types of questions are partial proxies for disadvantage and marginalization.<sup>335</sup> We should, therefore, expect to see higher “attitude” risk scores, on average, for people from historically marginalized backgrounds. In that sense, “attitude” responses may serve as racial proxies to a similar extent as other proxy factors.

There is ample empirical support for the proposition that the kinds of attitudes measured by risk assessment instruments are partial proxies for race. Meta-analyses and focused studies on risk assessment instruments and criminal thinking inventories have found that attitude scales have lower predictive accuracy for nonwhite populations but assign them higher risk scores.<sup>336</sup>

---

332. See, e.g., COMPAS Questionnaire, *supra* note 33, at 8; Idaho LSI-R Scoring Guide—Version 3.0, *supra* note 57, at 26 (requiring the assessor to evaluate the subject’s capacity for empathy and respect for rules).

333. COMPAS Questionnaire, *supra* note 33, at 8.

334. See *id.*

335. See Seth J. Prins, *Criminogenic or Criminalized? Testing an Assumption for Expanding Criminogenic Risk Assessment*, 43 LAW & HUM. BEHAV. 477, 479 (2019) (explaining that antisocial personality disorder and conduct disorder, which have “considerable overlap” with the “Big Four criminogenic risk factors,” are “structured by social disadvantage”).

336. See Mark E. Olver, Keira C. Stockdale & J. Stephen Wormith, *Thirty Years of Research on the Level of Service Scales: A Meta-Analytic Examination of Predictive Accuracy and Sources of Variability*, 26 PSYCH. ASSESSMENT 156, 162, 169-70 (2014) (finding that for nonwhite subjects, LS scores in most domains, including attitudes, were higher but had less predictive validity); Wilson & Gutierrez, *supra* note 188, at 207 (finding that the LSI-R has lower predictive ability for indigenous than non-indigenous subjects); Glenn D. Walters & Thomas H. Cohen, *Criminal Thought Process as a Dynamic Risk Factor: Variable- and Person-Oriented Approaches to Recidivism Prediction*, 40 LAW & HUM. BEHAV. 411, 414 (2016) (finding a positive, statistically significant relationship between race (nonwhite vs. white) and Psychological Inventory of Criminal Thinking Styles score for men, although not for women); June Price Tangney, Jeffrey Stuewig, Emi Furukawa, Sarah Kopelovich, Patrick J. Meyer & Brandon Cosby, *Reliability, Validity, and Predictive Utility of the 25-Item Criminogenic Cognitions Scale (CCS)*, 39 CRIM. JUST. & FOOTNOTE CONTINUED ON NEXT PAGE

Public opinion research also supports the conclusion that attitudes are correlated with race. In one 2018 survey, the Pew Research Center estimated “people’s general trust or distrust in others, their sense of the exploitative tendencies or fairness of others, and their assessment of the overall helpfulness or selfishness of others.”<sup>337</sup> The survey found higher rates of distrust among Black and Hispanic respondents (44% and 46% respectively were “low trusters” while 13% and 12% respectively were “high trusters”) than among white respondents (31% were “low trusters” and 27% were “high trusters”).<sup>338</sup> The NCSC’s 2017 survey on the state of state courts found a marked difference between Black and white respondents’ trust in judges.<sup>339</sup> When presented with contrasting statements about whether state court judges were in touch with the communities they served, 38% of white respondents, but only 19% of Black respondents, agreed more that judges were in touch with their communities, while 78% of Black respondents, but only 58% of white respondents, agreed more that judges were out of touch with their communities.<sup>340</sup> In a 2019 Pew Research Center survey, 44% of Black respondents said they had been unfairly stopped by the police, compared to 19% of Hispanic respondents, 16% of Asian respondents, and 9% of white respondents.<sup>341</sup> On such facts, one would expect to see higher criminal risk scores, on average, in marginalized populations because those risk scores include assessment of distrust toward civic institutions, authority, law enforcement, and the legal system.

### 3. Attitude measurement problems

Measuring attitudes to predict recidivism or assess “treatment needs” is a *very* inexact science. If there was strong agreement among developers on what a “criminal attitude” is and how to measure it, we would expect to see instruments achieve similar results in the attitude domain. That is not the case, at least for

---

BEHAV. 1340, 1356 (2012) (“Regarding race, African Americans scored higher than Whites on the total CCS scale as well as the dimensions of entitlement, insensitivity to the impact of crime, and negative attitudes toward authority.”).

337. Lee Rainie, Scott Keeter & Andrew Perrin, *Trust and Distrust in America*, PEW RSCH. CTR. (July 22, 2019), <https://perma.cc/R87N-68VD>.

338. *Id.*

339. GBA STRATEGIES, *supra* note 314, at 3.

340. *Id.* (showing what percentage of Black and white respondents agreed more with each of two statements: (a) “In general, judges in [my state’s] courts reflect the values of our communities and understand the challenges facing the people who appear in their courtrooms,” or (b) “Too many judges in [my state’s] courts don’t understand the challenges facing people who appear in their courtrooms and need to do a better job of getting out into the community and listening to people”).

341. JULIANA MENASCE HOROWITZ, ANNA BROWN & KIANA COX, PEW RSCH. CTR., RACE IN AMERICA 2019, at 42 (2019), <https://perma.cc/W3CA-P9J2>.

LSI-R and COMPAS. A 2010 study by UCLA researchers found that “COMPAS identified 37% of the sample as ‘high’ on criminal thinking, compared to 9% according to the LSI-R.”<sup>342</sup> The study also found that fewer than half of subjects who scored high on one scale also scored high on the other scale.<sup>343</sup>

One could conclude that one instrument may simply be better than the other at gauging criminal attitudes, but, in reality, neither appears to be very good at it. Two large-scale COMPAS validation studies found an AUC of 0.57 for the “Criminal Thinking” scale, which is little better than chance.<sup>344</sup> A 2020 validation of the LSI-R by the state of Hawaii found the same AUC of 0.57 for “Attitude Orientation.”<sup>345</sup> The Caudy study found AUCs of 0.50 and 0.58 for that same section in its two samples.<sup>346</sup>

Even when administering the same instrument, assessors tend to differ a lot on their scoring. A study of ORAS-CST inter-rater reliability by the Ohio Department of Rehabilitation and Correction found four out of seven items in the “Criminal Attitudes and Behavioral Patterns” to be “[u]nreliable” because inter-rater agreement was below 80%.<sup>347</sup> A study of LSI-R inter-rater reliability found 63% agreement on the “Supportive of crime” subsection and 72% agreement on the “Unfavorable toward convention” subsection.<sup>348</sup> There

---

342. FARABEE ET AL., *supra* note 153, at 20.

343. *Id.*

344. See COMPAS Guide, *supra* note 4, at 13, 23-24.

345. TIMOTHY WONG, INTERAGENCY COUNCIL ON INTERMEDIATE SANCTIONS, VALIDATION OF THE LSI-R AND ASUS CRIMINOGENIC RISK ASSESSMENT INSTRUMENTS: STATE OF HAWAII 2008–2016 REPORTING PERIOD 21 (2020), <https://perma.cc/E9QV-8AN7>.

346. Caudy et al., *supra* note 188, at 462.

347. GARY C. MOHR & SARA ANDREWS, DESCRIPTIVE AND RELIABILITY ANALYSIS FOR THE COMMUNITY SUPERVISION TOOL (ORAS-CST) 4 & n.# (2014); see CASEY ET AL., *supra* note 2, at 13 (explaining that inter-rater reliability, also called inter-rater agreement, “determines the degree to which different test administrators give the same offender similar scores on individual items as well as for the tool overall”).

348. AUSTIN ET AL., *supra* note 58, at 14 tbl.7. Leading criminologists found agreement percentages north of 90% on LSI-R attitude and associational subsections in a 2004 study. Christopher T. Lowenkamp, Alexander M. Holsinger, Lori Brusman-Lovins & Edward J. Latessa, *Assessing the Inter-rater Agreement of the Level of Service Inventory Revised*, FED. PROB., Dec. 2004, 34, 37. However, they stacked the decks by conducting the study at the end of a three-day professional training using printed vignettes instead of the open-ended interviews used in practice. See *id.* at 36 (“The participants in this training were part of a three-day [LSI-R scoring] training . . . . At the end of the training, the participants were given an exam that included a vignette describing an offender.”); *id.* at 37 (“[T]he LSI-R process, when conducted in the field, requires practitioners to gather their own data via one-on-one interviews with offenders and the consideration of multiple sources of collateral information. The subjects in the current study were given a tailor-made vignette that represented the information that should have been gathered had they been involved in a real-life assessment process. A true test of inter-rater reliability using the LSI-R would require pairs (or more) of  
*footnote continued on next page*

may be a consistent, reliable, and empirically valid way to measure attitudes that are correlated with recidivism, but the instruments discussed in this Article have not achieved it.

It is likely that the most salient “criminal attitude” questions are the least problematic from a constitutional standpoint and that the most problematic questions do not improve predictive validity. Intuitively, to predict how someone will act in the future, the way they treat others should be more consequential than *Dawson*-prohibited opinions on social issues. Some research supports that proposition, finding that “[b]ehavioral intentions are a better predictor of future behavior than are general attitudes.”<sup>349</sup> The COMPAS Questionnaire includes in the same section statements such as “I won’t hesitate to hit or threaten people if they have done something to hurt my friends or family” and “The law doesn’t help average people.”<sup>350</sup> The first statement is a description of past or potential behaviors, not a general attitude. The second statement is purely about general attitude. It is no stretch to presume that an attitude scale could improve its predictive power by dispensing with items that gauge general attitudes.

Constitutional, normative, and empirical considerations all counsel against current methods of quantifying attitudes. If an attitude scale could satisfy *Dawson*, achieve acceptable predictive utility, eliminate racial disparities, and facilitate high rates of inter-rater agreement, then quantification of attitudes might be useful to the criminal justice system. That is a tall order, and the instruments discussed in this Article do not satisfy it.

#### IV. Peer and Family Relationships

Risk assessment instruments include peer and family association items because of the social learning theory of crime, which posits that the “stability of criminal behavior is . . . more likely when an individual is embedded in a social environment where misconduct is reinforced and where differential association with pro-criminal definitions and behavioral patterns is readily available.”<sup>351</sup> That theory is intuitive but surprisingly hard to prove or translate into practical use.

---

subjects to gather their own information independently from the same source, after which the assessment would be scored accordingly.”).

349. Jeremy F. Mills, Daryl G. Kroner & Toni Hemmati, *The Measures of Criminal Attitudes and Associates (MCAA): The Prediction of General and Violent Recidivism*, 31 CRIM. JUST. & BEHAV. 717, 721 (2004) (citations omitted).

350. COMPAS Questionnaire, *supra* note 33, at 8.

351. Travis C. Pratt et al., *The Empirical Status of Social Learning Theory: A Meta-Analysis*, 27 JUST. Q. 765, 769 (2010).

The empirical research on “criminal peers” is limited by the difficulty of classifying and quantifying personal relationships and dogged by questions of causality.<sup>352</sup> For those reasons, it is difficult to determine whether associational factors are causal, particularly important, lawful, or ethically justified. It is possible, though, to draw some conclusions about the constitutionality of particular instruments and evaluate their design strengths and flaws.

There are powerful legal arguments against inclusion of family criminal history in criminal risk assessment and viable (although less powerful) arguments against inclusion of peer association items, as well as strong policy arguments against all associational and affiliation items. Instruments employ a wide variety of methods to quantify “criminal family” and “criminal acquaintances,” many of which are unsound. Because quantification requires reductive questions, risk assessment is a game played with blunt instruments. Personal relationships are nuanced, though—so nuanced that constitutional doctrine on private associations is comprised of unworkably vague and malleable standards. The nuances of human relationships should be captured through narrative, not quantification, in order to satisfy constitutional constraints and make sound, practical use of associational information.

#### A. Risk Assessment Associational Items

##### 1. Criminal peers

Risk assessment instruments commonly inquire into a subject’s criminal friends and acquaintances. ORAS-CST, IRAS-CST, and TRAS-CST require the assessor to determine whether some, none, or a majority of the subject’s friends are “criminal friends,” as well as how much contact the subject has with those friends on a scale of zero to two.<sup>353</sup> LSI-R assessors assign scores for “Some criminal acquaintances,” “Some criminal friends,” “Few anti-criminal acquaintances,” and “Few anti-criminal friends.”<sup>354</sup> The COMPAS questionnaire instructs the subject to “think of your friends and the people you hung out with in the past few (3-6) months” and then answer “None,” “Few,” “Half,” or “Most” to four questions: “How many of your friends/acquaintances have ever been

---

352. See *infra* Part IV.D.

353. COLLINS ET AL., *supra* note 71, app. E at 60 (ORAS-CST); ELEK ET AL., *supra* note 25, app. at A-79 (IRAS-CST); UNIV. OF CINCINNATI & TEX. DEP’T OF CRIM. JUST., *Texas Risk Assessment System—Community Supervision—Felony Score Sheet*, in TEXAS RISK ASSESSMENT SYSTEM PARTICIPANT MANUAL, *supra* note 71, at 3 [hereinafter Texas Felony Score Sheet]; UNIV. OF CINCINNATI & TEX. DEP’T OF CRIM. JUST., *Texas Risk Assessment System—Community Supervision—Misdemeanor Score Sheet*, in TEXAS RISK ASSESSMENT SYSTEM PARTICIPANT MANUAL, *supra* note 71, at 2 [hereinafter Texas Misdemeanor Score Sheet].

354. Idaho LSI-R Scoring Guide—Version 3.0, *supra* note 57, at 16-17.

arrested?"; "How many of your friends/acquaintances served time in jail or prison?"; "How many of your friends/acquaintances are gang members?"; and "How many of your friends/acquaintances are taking illegal drugs regularly (more than a couple times a month)?"<sup>355</sup> The OST instructs the subject to characterize their "current peer and/or associate relationships" as either "Positive, law-abiding" or "Negative, criminally oriented."<sup>356</sup> All six instruments also inquire into the level of criminal activity in the subject's neighborhood.<sup>357</sup>

COMPAS factors into its assessment not only whether the subject's acquaintances have *committed* crimes, but also whether the subject's neighborhood friends or family have been crime *victims*.<sup>358</sup> This element is indefensible. The idea that one should present as higher risk for choosing to associate with people who have committed crimes is arguably reasonable, but the idea that one should present as higher risk for choosing to associate with victims is morally repugnant.

## 2. Family criminal history

Factoring family criminal history into risk scores is effectively kin punishment—an archaic-sounding term for an atavistic practice. The Code of Hammurabi mandates that if a home collapses due to faulty construction and kills the owner's son, the builder's son will be put to death.<sup>359</sup> In Greek legend, Apollo and Artemis kill Niobe's children to punish her for insulting their mother.<sup>360</sup> In the Old Testament, God slays every Egyptian firstborn son and threatens to visit the iniquities of the Israelite parents who hate him upon their descendants down to the fourth generation.<sup>361</sup> Contemporary risk assessment also penalizes individuals for the crimes of their parents. And their siblings. And their spouses.

Many states that bar consideration in penalty decisions of characteristics that are typically (although not always) immutable, such as race, gender, religion, social status, and wealth, fail to similarly bar consideration of family

---

355. COMPAS Questionnaire, *supra* note 33, at 3.

356. STINSON, *supra* note 81, app. 3 at 88.

357. COLLINS ET AL., *supra* note 71, app. E at 59 (ORAS-CST); ELEK ET AL., *supra* note 25, app. at A-78 (IRAS-CST); Texas Felony Score Sheet, *supra* note 354, at 2; Idaho LSI-R Scoring Guide—Version 3.0, *supra* note 57, at 15; COMPAS Questionnaire, *supra* note 33, at 4-5; STINSON, *supra* note 81, app. 3 at 88-89 (OST).

358. COMPAS Questionnaire, *supra* note 33, at 5.

359. THE CODE OF HAMMURABI, KING OF BABYLON § 230, at 81 (Robert Francis Harper trans., 2d ed. 1904) (c. 2250 B.C.E.).

360. HOMER, THE ILLIAD 416 (Samuel Butler trans., New York, Dutton & Co. 1898) (c. 800 B.C.E.).

361. *Exodus* 12:29, 20:5.



*What Even Is a Criminal Attitude?*  
75 STAN. L. REV. 1431 (2023)

history.<sup>362</sup> Numerous risk assessment instruments licensed or developed by states include family members' actions beyond the defendant's control as factors. The tables below detail the family criminality questions and items for COMPAS, the LSI-R, the OST, and ORAS-CST.

**Table 9**  
Comparison of Family Criminality Questions

| <b>COMPAS: "Family Criminality"</b> <sup>363</sup>  | <b>Idaho LSI-R Scoring Guide: "Criminal Family/Spouse"</b> <sup>364</sup>  | <b>West Virginia LSI-R Questionnaire: "Criminal Involvement"</b> <sup>365</sup>   |
|---|--|---|
| <p>Which of the following best describes who principally raised you?</p> <p>If you lived with both parents and they later separated, how old were you at the time?</p> <p>Was your father (or father figure who principally raised you) ever arrested, that you know of?</p> <p>Was your mother (or mother figure who principally raised you) ever arrested, that you know of?</p> <p>Were your brothers or sisters ever arrested, that you know of?</p> <p>Was your wife/husband/partner ever arrested, that you know of?</p> <p>Did a parent or parent figure who raised you ever have a drug or alcohol problem?</p> <p>Was one of your parents (or parent figure who raised you) ever sent to jail or prison?</p> | <p><i>Criminal family member(s) indicate negative modeling and exposure to pro-criminal influences and/or vicarious reinforcement of anti-social attitudes and behaviors.</i></p> <p>(a) Score "Yes" if an offender's parent, sibling, current spouse/equivalent, or other <u>close</u> relative has a criminal record, or an established pattern of criminal behavior.</p> <p>(b) Criminal record is defined as any conviction, deferred judgments, or deferred prosecution for any criminal misdemeanor or felony.</p> | <p>(1) Does anyone in your family have a criminal record? Who?</p> <p>(2) Has either of your parents, spouse or significant other engaged in anti-social/criminal activity? Explain.</p> <p>(3) At what point in your life were they involved in the criminal activity?</p> <p>(4) How much of an influence have they had on your life?</p> <p>(5) What do you think of their behavior? How do you feel about them?</p> |

362. See Hamilton, *supra* note 7, at 274 & n.320 (listing explicitly prohibited sentencing factors in several states, none of which include family history).

363. COMPAS Questionnaire, *supra* note 33, at 3.

364. Idaho LSI-R Scoring Guide—Version 3.0, *supra* note 57, at 14.

365. COLLINS ET AL., *supra* note 71, app. D at 54-55.

| <b>OST Scoresheet: “Family and Social Relationships”<sup>366</sup></b>  | <b>ORAS-CST: “Family and Social Support” Parent Criminality Item<sup>367</sup></b> |
|---|--|
| <p>Was one of your parents (or parent figure who raised you) ever sent to jail or prison?</p> <p>Was the client raised primarily by a biological parent?</p> <p>Did the client live in the family home until age 16?</p> <p>Does anyone in the client’s family (parents or siblings) have a criminal record?</p> <p>What is the status of the current family relations: [Positive/Negative]</p> <p>What is the client’s childhood history of domestic violence? (As victim or witness)</p> <p>What best describes the client’s current most significant companion/romantic relationship?</p> <p>What best describes the client’s current peer and/or associate relationships?</p> <p>Does the client associate with at least one prosocial person on a regular basis?</p> | <p>Parents have Criminal Record</p> <p>0 = No</p> <p>1 = Yes</p>                   |

**B. Constitutional Protections for Peer and Family Relationships**

The analysis that follows assesses the strength of a theoretical challenge to peer and family association items under core substantive due process, First Amendment, and equal protection principles. It seeks some signal in the noise of intimate association doctrine for peer association items and looks to the closest analogous equal protection cases for the analysis of family affiliation items. Peer association items may violate the First Amendment in some instances although they probably do not violate substantive due process. The

---

366. STINSON, *supra* note 81, app. 3 at 99.

367. COLLINS ET AL., *supra* note 71, app. E at 59.

equal protection case against family affiliation items is less equivocal; quantification of family criminal history should be prohibited in all criminal justice proceedings.

1. The substantive due process dead end

a. The *Jaycees* quagmire

The Constitution protects two strains of freedom of association: expressive and intimate. The Supreme Court explained in *Roberts v. United States Jaycees*:

Our decisions have referred to constitutionally protected “freedom of association” in two distinct senses. In one line of decisions, the Court has concluded that choices to enter into and maintain certain intimate human relationships must be secured against undue intrusion by the State because of the role of such relationships in safeguarding the individual freedom that is central to our constitutional scheme. In this respect, freedom of association receives protection as a fundamental element of personal liberty. In another set of decisions, the Court has recognized a right to associate for the purpose of engaging in those activities protected by the First Amendment—speech, assembly, petition for the redress of grievances, and the exercise of religion. The Constitution guarantees freedom of association of this kind as an indispensable means of preserving other individual liberties.<sup>368</sup>

*Jaycees* lays out a spectrum of intimacy in private associations from “the most intimate to the most attenuated of personal attachments.”<sup>369</sup> At the “most intimate” pole lie associations that “attend the creation and sustenance of a family—marriage, childbirth, the raising and education of children, and cohabitation with one’s relatives.”<sup>370</sup> Those associations are “distinguished by such attributes as relative smallness, a high degree of selectivity in decisions to begin and maintain the affiliation, and seclusion from others in critical aspects of the relationship.”<sup>371</sup> At the “most attenuated” pole lie impersonal entities “such as a large business enterprise.”<sup>372</sup> “Between these poles, of course,” the Court noted, “lies a broad range of human relationships that may make greater or lesser claims to constitutional protection from particular incursions by the State.”<sup>373</sup>

The *Jaycees* Court declined to “mark the potentially significant points” on the intimacy spectrum “with any precision,” generating considerable

---

368. 468 U.S. 609, 617-18 (1984).

369. *Id.* at 620.

370. *Id.* at 619 (citations omitted).

371. *Id.* at 620.

372. *Id.*

373. *Id.*

ambiguity about what kinds of relationships are protected.<sup>374</sup> Freedom of intimate association is not limited to family per se, as *Jaycees* protects relationships with the qualities that close family relationships “exemplify.”<sup>375</sup> Lower courts, however, have had a difficult time demarcating what kinds of peer—or even family—associations are sufficiently intimate to warrant constitutional protection.<sup>376</sup>

Because *Jaycees* did not situate freedom of intimate association under a specific amendment, courts differ not only on its contours, but also on its constitutional source.<sup>377</sup> The Supreme Court itself has evinced internal disagreement about the right’s source in post-*Jaycees* intimate association jurisprudence.<sup>378</sup> In the absence of Supreme Court guidance, lower courts fall all over the map on whether freedom of intimate association is protected by the First Amendment, the Fourteenth Amendment, or both. Some courts assert that the governing standard remains the same no matter the source, some choose a lane, some acknowledge the ambiguity and decline to choose a lane, and some understand the First and Fourteenth Amendments to generate independent intimate association protections.<sup>379</sup> Intimate association presents a smorgasbord of potential analytical frameworks.

---

374. *Id.*

375. *Id.* at 619-20 (stating that only relationships with the “sorts of qualities” that close family relationships “exemplify” are likely to invoke constitutional protections).

376. To invoke but a few examples: The Third Circuit has held that friendships are categorically unprotected. *Martsolf v. Christie*, 522 F. App’x 149, 152 (3d Cir. 2013). The Sixth Circuit “has determined that a [p]ersonal friendship is protected as an intimate association.” *Medlin v. City of Algood*, 814 F. App’x 7, 18 (6th Cir. 2020) (quoting *Akers v. McGinnis*, 352 F.3d 1030, 1039-40 (6th Cir. 2003)). The Eleventh Circuit has held that “a dating relationship of about four years constitutes an intimate relationship” protected by the First Amendment but that a friendship without any “traits . . . inherent to family relationships” receives no constitutional protection. *Moore v. Tolbert*, 490 F. App’x 200, 203-04 (11th Cir. 2012) (per curiam).

377. *See Adler v. Pataki*, 185 F.3d 35, 42 (2d Cir. 1999) (observing that the “nature and extent” of the right of intimate association is “hardly clear”).

378. *See City of Dallas v. Stanglin*, 490 U.S. 19, 23-24 (1989) (discussing intimate association as a First Amendment right); *id.* at 28 (Stevens, J., concurring in the judgment) (situating the right under Fourteenth Amendment substantive due process); *Lyng v. Int’l Union, UAW*, 485 U.S. 360, 362, 365-66 (1988) (implicitly accepting the First Amendment as a source of authority for a challenge to a law that penalized food stamp-eligible households if a household member participated in a strike); *Lawrence v. Texas*, 539 U.S. 558, 578 (2003) (situating intimate association under substantive due process); *Obergefell v. Hodges*, 576 U.S. 644, 675-76 (2015) (same).

379. *See, e.g., Shahar v. Bowers*, 114 F.3d 1097, 1113 n.6 (11th Cir. 1997) (Tjoflat, J., concurring) (describing the circuit split on where the right of intimate association is situated); *Mann v. City of Sacramento*, 748 F. App’x 112, 115 (9th Cir. 2018) (mem.) (“Because we analyze the right of intimate association in the same manner regardless [of] whether we characterize it under the First or Fourteenth Amendments, *Ward* necessarily rejected any argument that adult, non-cohabitating siblings enjoy a right to

*footnote continued on next page*

b. Peer association items do not violate substantive due process

There are substantial constraints on due process protections for chosen relationships.<sup>380</sup> Only chosen *family* relationships—procreative and spousal—are unequivocally entitled to some measure of constitutional protection.<sup>381</sup> Even those relationships, though, are not protected from all government interference in all circumstances, and the type of relationship and government action determine which part of the Constitution protects those rights.<sup>382</sup> In the marriage context, government regulations that “interfere directly and substantially” with the decision to marry are subject to strict scrutiny, while regulations that “do not significantly interfere” are subject only to rational basis review.<sup>383</sup> Since risk assessment associational items cannot be said to interfere directly and substantially with the decision to marry, a legal challenge to those items based on a marital relationship would be subject to rational basis review. It stands to reason that if even the marital relationship warrants only rational basis review, less protected relationships would not warrant a higher level of scrutiny.<sup>384</sup>

The Due Process Clause is, therefore, not conducive to a constitutional challenge to risk assessment associational items. Rational basis review requires

---

intimate association.” (citing *Ward v. City of San Jose*, 967 F.2d 280, 283-84 (9th Cir. 1991)); *Mann v. Sacramento Police Dep’t*, 803 F. App’x 142, 143-44 (9th Cir. 2020) (stating that *Ward*, in fact, forecloses only a Fourteenth Amendment intimate association claim, not a First Amendment intimate association claim).

380. Family criminal history items are discussed in an equal protection framework in Part IV.B.3 below but are excluded from the substantive due process analysis because intimate association protections are designed to safeguard chosen affiliations, not unchosen ties. See *Jaycees*, 468 U.S. at 617-618 (describing freedom of intimate association as a constitutional protection for “choices to enter into and maintain certain intimate human relationships”); *Lawrence*, 539 U.S. at 575 (distinguishing between “[e]quality of treatment” under the Equal Protection Clause and “the due process right to demand respect for conduct protected by the substantive guarantee of liberty”).
381. See *Carey v. Population Servs. Int’l*, 431 U.S. 678, 684-85 (1977) (“While the outer limits of this aspect of privacy have not been marked by the Court, it is clear that among the decisions that an individual may make without unjustified government interference are personal decisions ‘relating to marriage, procreation, contraception, family relationships, and child rearing and education.’” (citations omitted) (quoting *Roe v. Wade*, 410 U.S. 113, 152-53 (1973))).
382. See *Adler*, 185 F.3d at 42-43 (searching for a common thread in cases that variously situate constitutional protections for the marital relationship under the First Amendment, Fourteenth Amendment substantive due process, and Fourteenth Amendment equal protection).
383. See *Zablocki v. Redhail*, 434 U.S. 374, 386-87 (1978).
384. The Sixth Circuit has expressly adopted this reasoning, applying heightened scrutiny to a policy that directly or substantially interferes with freedom of intimate association and rational basis review otherwise. See *Akers v. McGinnis*, 352 F.3d 1030, 1039-40 (6th Cir. 2003).

only that a government action be “rationally related to legitimate government interests.”<sup>385</sup> Inarguably, the government has a legitimate interest in crafting an appropriate criminal sanction,<sup>386</sup> and an inquiry into a defendant’s associates is rationally related to that goal. Even empirical proof that criminal peers do not increase recidivism risk would not change the picture, since a government action “unsupported by evidence or empirical data” survives rational basis review if it is “based on rational speculation.”<sup>387</sup>

Due process protections against association items are further weakened by the indirect nature of the relationship between risk score and penalty. A higher risk score does not necessarily result in adverse consequences for the defendant. Risk scores are never the sole factor a decisionmaker considers and they are subject to override if the decisionmaker disagrees with the score.<sup>388</sup> The combination of an attenuated link between factor and punishment and weak substantive due process protections for personal relationships renders substantive due process a dead end for a constitutional challenge to the use of associational factors in criminal risk assessment.

## 2. The First Amendment potential pathway

The First Amendment offers protections that the Fourteenth Amendment does not because government restrictions on First Amendment-protected speech and associations trigger strict scrutiny.<sup>389</sup> Government action that

---

385. *Box v. Planned Parenthood of Ind. & Ky., Inc.*, 139 S. Ct. 1780, 1781-82 (2019) (quoting *Washington v. Glucksberg*, 521 U.S. 702, 728 (1997)).

386. *See Schall v. Martin*, 467 U.S. 253, 264 (1984) (“The ‘legitimate and compelling state interest’ in protecting the community from crime cannot be doubted.” (quoting *De Veau v. Braisted*, 363 U.S. 144, 155 (1960))).

387. *FCC v. Beach Commc’ns*, 508 U.S. 307, 315 (1993).

388. *See, e.g., State v. Samsa*, 859 N.W.2d 149, 153 (Wis. Ct. App. 2014) (“And I think basically what happened here is that the Court rejected the . . . agent’s assessment and the COMPAS assessment that . . . Samsa was a low risk. The Court determined that under all of the circumstances that exist here he’s a higher risk than that.”); *Gage v. Richardson*, No. 16-cv-849, 2019 WL 1900338, at \*3 (W.D. Wis. Apr. 29, 2019) (“The judge noted that he was not giving the [COMPAS] assessment much weight because he thought Gage’s continued denial of wrongdoing and his moving to Canada without notifying the court showed that Gage needed extensive supervision.”); *see also Garrett & Monahan, supra* note 176, at 464-65 (discussing broad variation in how Virginia judges sentenced defendants with the same risk score and citing a four-state survey of judges that found overrides were common in juvenile cases).

389. *See Reed v. Town of Gilbert*, 135 S. Ct. 2218, 2226-27 (2015) (explaining that government regulation of speech is subject to strict scrutiny when it “draws distinctions based on the message a speaker conveys”); *Roberts v. U.S. Jaycees*, 468 U.S. 609, 623 (1984) (“Infringements on [expressive association] may be justified by regulations adopted to serve compelling state interests, unrelated to the suppression of ideas, that cannot be achieved through means significantly less restrictive of associational freedoms.”).

infringes on First Amendment rights is constitutional only “if it furthers an important or substantial governmental interest; if the governmental interest is unrelated to the suppression of free expression; and if the incidental restriction on alleged First Amendment freedoms is no greater than is essential to the furtherance of that interest.”<sup>390</sup>

Peer association items meet the first two criteria, but fail on the third. The goal of peer association items can be more narrowly achieved by reliance on narrative than quantification.<sup>391</sup> Instruments that factor peer associations into the risk score do not outperform instruments that rely on a small number of static factors.<sup>392</sup> For instruments that do not factor peer associations into the risk score but generate a discrete peer association need score, the government’s interests are better served through narrative than scoring.<sup>393</sup> A judge needs to know the qualities of the defendant’s relationships, not a number, to make an individualized determination.

Additionally, peer association items can infringe on the freedom of expressive association. Instruments will penalize a subject who frequently associates with “criminal peers” for the purpose of making music or attending protests to the same degree as a subject who associates with “criminal peers” for purely social purposes.<sup>394</sup> The possibility that expressive associations increase some subjects’ risk scores increases the viability of a challenge by reference to *Dawson* even in jurisdictions that situate freedom of intimate association under the Fourteenth Amendment.

The First Amendment also offers greater protection than the Fourteenth because of the overbreadth doctrine. Because “First Amendment freedoms need breathing space to survive,” courts safeguard those freedoms by protecting against government restrictions even at the margins.<sup>395</sup> In a First Amendment context, “even though the governmental purpose be legitimate and substantial, that purpose cannot be pursued by means that broadly stifle fundamental personal liberties when the end can be more narrowly achieved.”<sup>396</sup> For that reason, the First Amendment overbreadth doctrine permits “attacks on overly broad statutes with no requirement that the person making the attack

---

390. *United States v. O’Brien*, 391 U.S. 367, 377 (1968).

391. *See supra* Part II.C.

392. *See supra* Part II.C.

393. *See supra* Part II.C.

394. *See, e.g.*, Idaho LSI-R Scoring Guide—Version 3.0, *supra* note 57, at 16-17 (drawing no distinctions based on activity type); TRAS-CST Interview Guide, *supra* note 71, at 2-17, 2-34 (scoring “Criminal Friends” and “Contact with Criminal Peers” items with no reference to activity type).

395. *NAACP v. Button*, 371 U.S. 415, 433 (1963).

396. *Shelton v. Tucker*, 364 U.S. 479, 488 (1960).

demonstrate that his own conduct could not be regulated by a statute drawn with the requisite narrow specificity.”<sup>397</sup>

Because there is no Fourteenth Amendment overbreadth doctrine, courts that have considered overbreadth challenges to statutes and regulations restricting social relationships have held against the plaintiff if they understood the Fourteenth Amendment to be the source of freedom of intimate association.<sup>398</sup> For example, in a 2020 decision, a federal district judge in Michigan analyzed an intimate association claim under the Fourteenth Amendment, pursuant to Sixth Circuit precedent, and found that a regulation that barred sheriff’s deputies from regular social contact with felons survived rational basis review even though it was unfairly broad.<sup>399</sup> By way of contrast, there have been at least two successful overbreadth challenges based on freedom of intimate association when a court situated the right under the First Amendment.<sup>400</sup>

Peer association risk assessment domains certainly sweep broadly, capturing common scenarios in which associates are unlikely to have a negative influence. A defendant could receive peer association points for having friends who were arrested but never convicted, or a friend who committed a felony twenty years ago and has been law-abiding ever since.<sup>401</sup>

---

397. *Dombrowski v. Pfister*, 380 U.S. 479, 486 (1965).

398. See *Bailey v. City of Nat’l City*, 277 Cal. Rptr. 427, 433-34 (Ct. App. 1991); *Frazier v. State*, 2016 WL 4437996, at \*2-3 (Nev. Aug. 10, 2016); *Hvamstad v. Suhler*, 727 F. Supp. 511, 517 (D. Minn. 1989) (“The overbreadth doctrine, in its limited state, cannot however be extended to encompass intimate association rights that are based not on first amendment concepts, but on general liberty notions.”), *aff’d*, 915 F.2d 1218 (8th Cir. 1990) (per curiam).

399. *Duncan v. Cnty. of Wayne*, No. 19-12211, 2020 WL 6273452, at \*4-5 (E.D. Mich. Oct. 26, 2020).

400. *Sawyer v. Sandstrom*, 615 F.2d 311, 315-16 (5th Cir. 1980) (holding a loitering ordinance unconstitutional); *State v. Holiday*, 585 N.W.2d 68, 71 n.1 (Minn. Ct. App. 1998) (“There is some dispute whether the freedom of intimate association involved here, that is unrelated to expressive conduct protected by the First Amendment, is subject to the overbreadth doctrine. . . . We believe, however, that it is appropriate to apply the overbreadth doctrine to such a sweeping limitation on the freedom of association.”). *But see United States v. Estrada-Tepal*, 57 F. Supp. 3d 164, 170-71, 170 n.6 (E.D.N.Y. 2014) (entertaining, but ultimately rejecting, a facial overbreadth challenge to a federal sex trafficking statute without committing to a First Amendment locus).

401. See, e.g., UNIV. OF CINCINNATI CTR. FOR CRIM. JUST. RSCH., OHIO RISK ASSESSMENT SYSTEM—COMMUNITY SUPERVISION TOOL (ORAS-CST) SCORING GUIDE, in 2 NEV. JUD. COUNCIL, *supra* note 62, at 2-9, 2-17; UNIV. OF CINCINNATI CTR. FOR CRIM. JUST. RSCH., *supra* note 62, at 2-34 (not distinguishing between friends and peers with criminal histories and those actively engaging in criminal behavior).



Defendants can also receive peer association points if they rely on people with criminal histories for childcare, housing, or transportation.<sup>402</sup>

Overbreadth is a high bar, though, a “strong medicine” that “has been employed by the [Supreme] Court sparingly and only as a last resort.”<sup>403</sup> Moreover, when the challenged law or regulation pertains to conduct, the doctrine’s function “attenuates.”<sup>404</sup> Thus, the standard when conduct is involved is that the overbreadth “must not only be real, but substantial as well, judged in relation to the statute’s plainly legitimate sweep.”<sup>405</sup>

The consequences of peer association item overbreadth may not be substantial. Risk-needs scores do not criminalize conduct, nor do they directly penalize a defendant for their associations. Rather, peer association scores increase, to some indeterminate degree, the likelihood of penalties like a heavier sentence or additional probation conditions. Moreover, some instruments attempt to distinguish between criminogenic and innocent associations, albeit with mixed success.<sup>406</sup>

In sum, in First Amendment intimate association jurisdictions, peer association items are likely unconstitutional as applied to specific defendants and possibly susceptible to attack under the overbreadth doctrine. Risk assessment items with no true utility may survive rational basis review under the Fourteenth Amendment, but they should not survive strict scrutiny under the First Amendment, because peer association items do not improve the predictive validity of the instrument and are less useful than narrative explanations. In jurisdictions that situate freedom of intimate association under the First Amendment, defendants who score points on peer association items could make a viable claim that those items violate their constitutional rights.

### 3. The Equal Protection proscription of kin punishment

Individual responsibility is a core tenet of the American justice system. The Supreme Court has been emphatic on that point: “Our law punishes people for what they do, not who they are. Dispensing punishment on the basis of an

---

402. *See id.* (making no exception for people who rely on friends or peers with criminal histories out of necessity).

403. *Broadrick v. Oklahoma*, 413 U.S. 601, 613 (1973).

404. *Id.* at 615.

405. *Id.*

406. *See, e.g., Idaho LSI-R Scoring Guide—Version 3.0, supra* note 57, at 16 (“For acquaintances or friends that have criminal records but are now clearly pro-social and stable e.g., AA/NA sponsor with several years clean and sober, do not count these individuals as a pro-criminal influence.”). *But see id.* at 17 (omitting similar admonitions for other items in the “Companions” category).

immutable characteristic flatly contravenes this guiding principle.”<sup>407</sup> Equal protection jurisprudence, in particular, enshrines the precept that “legal burdens should bear some relationship to individual responsibility or wrongdoing.”<sup>408</sup> Because kin punishment contravenes the fundamental principle of individual responsibility, it is vulnerable to attack on equal protection grounds.<sup>409</sup>

Two scholars have applied equal protection principles to “criminal family” factors by way of analogy to illegitimacy and immigration status. Dawinder Sidhu has referenced the 1982 immigration status decision *Plyler v. Doe* to support a normative argument against kin punishment in risk assessment.<sup>410</sup> Sonja Starr has made the normative argument that “to sentence people based on their family members’ or neighbors’ conduct runs afoul of principles of individual responsibility and adds the state’s imprimatur to the stigma people from troubled family backgrounds already face.”<sup>411</sup> She is skeptical, though, about the viability of a legal challenge to “sentencing based on a parent’s criminal record,” writing that it “could rely on a very strong analogy to illegitimacy” but would “push the boundaries of current doctrine.”<sup>412</sup> A challenge premised on analogy to both illegitimacy and immigration status, however, may be less boundary-pushing. Drawing on illegitimacy jurisprudence and *Plyler*, one could make a highly plausible case for the application of intermediate scrutiny to risk assessment questions about family members’ actions.

One can only argue by analogy because it is so obvious that no one should be punished for their parent’s crimes, criminal case law rarely speaks to the question. While there is a substantial body of case law on “guilt by association,” those precedents address only convictions, not pretrial or postconviction penalties based on associations.<sup>413</sup> There is, however, relevant civil

---

407. *Buck v. Davis*, 137 S. Ct. 759, 778 (2017).

408. *Weber v. Aetna Cas. & Sur. Co.*, 406 U.S. 164, 175 (1972).

409. *See Plyler v. Doe*, 457 U.S. 202, 218 n.16 (1982) (“Only when concerns sufficiently absolute and enduring can be clearly ascertained from the Constitution and our cases do we employ [intermediate scrutiny] to aid us in determining the rationality of the legislative choice.”); *Weber*, 406 U.S. at 172-73 (“Though the latitude given state economic and social regulation is necessarily broad, when state statutory classifications approach sensitive and fundamental personal rights, this Court exercises a stricter scrutiny.” (citing *Brown v. Bd. of Educ.*, 347 U.S. 483 (1954); and *Harper v. Va. Bd. of Elections*, 383 U.S. 663 (1966))).

410. Sidhu, *supra* note 16, at 719.

411. Starr, *supra* note 296, at 230.

412. *Id.* at 231.

413. *See, e.g., Scales v. United States*, 367 U.S. 203, 224-27 (1961) (discussing whether organizational membership can support a criminal conviction); *United States v. Lane*, 474 U.S. 438, 462 (1986) (Brennan, J., concurring in part and dissenting in part)

*footnote continued on next page*

jurisprudence on non-criminal associational penalties. In *Plyler v. Doe*, the Supreme Court struck down a law that denied public school education to the children of undocumented immigrants.<sup>414</sup> Justice Brennan, writing for the majority, expressed that “legislation directing the onus of a parent’s misconduct against his children does not comport with fundamental conceptions of justice.”<sup>415</sup> Similarly, the Court asserted in *Weber v. Aetna Casualty & Surety Co.* that “imposing disabilities on the illegitimate child is contrary to the basic concept of our system that legal burdens should bear some relationship to individual responsibility or wrongdoing.”<sup>416</sup> These statements are dicta, but they are powerful dicta, expressing notions about “fundamental conceptions of justice”<sup>417</sup> and a “basic concept of our system.”<sup>418</sup> Kin punishment is just as antithetical to fundamental conceptions of justice in a criminal risk assessment context as it is in an immigration or insurance context.

Whereas the constitutional standards relevant to freedom of speech and association vary at different phases of the criminal justice process, kin punishment should be unconstitutional at every stage under *Plyler* and *Weber*. The two cases support the premise that unconstitutional “disabilities” for family behavior are not limited to denial of rights. In *Plyler*, the disability was undocumented children’s ineligibility for free public education and the Court emphasized that public education is not a constitutional right.<sup>419</sup> In *Weber*, the disability was the priority of legitimate children over illegitimate children in allocating a deceased parent’s worker’s compensation, which is also not a constitutional entitlement.<sup>420</sup> Pretrial incarceration, imposition of prison time instead of probation, and denial of parole are undeniably serious disabilities as

---

(citing *McElroy v. United States*, 164 U.S. 76, 80 (1896)) (explaining that the rules governing joinder are intended to prevent jury “inferences of habitual criminality or guilt by association”).

414. 457 U.S. at 230 (“If the State is to deny a discrete group of innocent children the free public education that it offers to other children residing within its borders, that denial must be justified by a showing that it furthers some substantial state interest. No such showing was made here.”).

415. *Id.* at 220.

416. 406 U.S. 164, 175 (1972). The situation in *Weber* was somewhat different, as the case involved a law meant to deter parental conduct that the court concluded served no state interest whatsoever. Use of predictive algorithms does serve a state interest and is meant to punish the defendant, not the defendant’s parents. Nonetheless, the principle that no one should be punished for their family members’ actions should prevail in any constitutional balancing test.

417. *Plyler*, 457 U.S. at 220.

418. *Weber*, 406 U.S. at 175.

419. *Plyler*, 457 U.S. at 221 (“Public education is not a ‘right’ granted to individuals by the Constitution.” (citing *San Antonio Indep. Sch. Dist. v. Rodriguez*, 411 U.S. 1, 35 (1973))).

420. *Weber*, 406 U.S. at 165.

well, as they amount to denial of freedom. The principles delineated in *Plyler* and *Weber* should apply to the criminal justice context because the law should not countenance any disability for family behavior in any context.

Kin punishment entails invidious discrimination, which occurs “[w]hen the law lays an unequal hand on those who have committed precisely the same offense.”<sup>421</sup> A state action is “presumptively invidious” if it “disadvantage[s] a ‘suspect class,’ or . . . impinge[s] on the exercise of a ‘fundamental right.’”<sup>422</sup> A right is fundamental when it is “so rooted” in the “traditions and collective conscience of our people” that it “cannot be denied without violating those ‘fundamental principles of liberty and justice which lie at the base of all our civil and political institutions.’”<sup>423</sup> The Supreme Court’s powerful dicta on individual liability—that it is a “basic concept of our system” and that filial liability offends “fundamental conceptions of justice”—should suffice to establish freedom from kin punishment as a fundamental right.<sup>424</sup>

Because family criminal history items in risk assessment constitute invidious discrimination, an equal protection challenge would be subject to intermediate scrutiny, the standard applied in *Plyler* and *Weber*, under which the discrimination at issue is constitutional only if it furthers a substantial goal of the state.<sup>425</sup> Justice Powell, writing for the majority in *Weber*, framed intermediate scrutiny as a balancing test: “The essential inquiry . . . [is] inevitably a dual one: What legitimate state interest does the classification promote? What fundamental personal rights might the classification endanger?”<sup>426</sup>

Family criminal history items violate equal protection because they infringe on a fundamental right while failing to promote government interests. The government’s interest in crafting an appropriate, individualized sanction is certainly legitimate, but classification on the basis of family criminality does not promote that interest for either risk or need assessment purposes. Quantification of family characteristics does not improve predictive power; in fact, narrative on family background offers richer, more precise information than numbers. The government interest side of the balancing test is, therefore, weak. On the fundamental rights side, however, lies the “basic concept of our system” that legal penalties should be based on individual

---

421. *Selective Serv. Sys. v. Minn. Pub. Int. Rsch. Grp.*, 468 U.S. 841, 880 (1984) (Marshall, J., dissenting) (citing *Skinner v. Oklahoma ex rel. Williamson*, 316 U.S. 535, 541 (1942)).

422. *Plyler*, 457 U.S. at 216-17.

423. *Griswold v. Connecticut*, 381 U.S. 479, 493 (1965) (Goldberg, J., concurring) (first quoting *Snyder v. Massachusetts*, 291 U.S. 97, 105 (1934); and then quoting *Powell v. Alabama*, 287 U.S. 45, 67 (1932)).

424. See *Weber*, 406 U.S. at 175; *Plyler*, 457 U.S. at 220.

425. *Plyler*, 457 U.S. at 224.

426. *Weber*, 406 U.S. at 173.

culpability.<sup>427</sup> Equal protection does not countenance quantification of family criminal history.

### C. Echoes of Eugenics in Criminal Risk Assessment

Eugenics had an overt influence on early criminal risk assessment research. Ernest Burgess, the Chicago School godfather of criminal risk assessment, included an inmate's father's race or nationality among the variables he studied.<sup>428</sup> Burgess's coauthor speculated in their seminal 1925 book, *The City*, that people with innate antisocial tendencies tend to group together in cities, finding "moral support for the traits they have in common which they would not find in a less select society."<sup>429</sup> He believed that nationality was the likely root of such tendencies, writing, "it has often occurred to me that those long genealogies of the Jukes and the tribes of Ishmael would not show such a persistent and distressing uniformity of vice, crime, and poverty unless they were peculiarly fit for the environment in which they are condemned to exist."<sup>430</sup> Burgess's student, Clark Tibbitts, wrote in 1931 that the relevance of the father's nationality to risk assessment is so "obvious," it "require[s] little comment."<sup>431</sup>

To this day, there are eugenic echoes in risk assessment literature. J.C. Oleson noted in 2011 that "[e]arly theorists believed that crime ran in deviant families" because "[a]fter all, it is said that 'the acorn does not fall far from the tree.'"<sup>432</sup> He threaded the needle, however, on the question of genetic influence, writing that "while there are thorny and unanswered questions about the relative contributions of environmental, biological, psychological, genetic, and social influences on crime, research consistently indicates that criminal parents are more likely to raise criminal children than non-criminal parents."<sup>433</sup> COMPAS's official materials are more direct: "Aside from the social learning and role modeling perspective," the COMPAS Guide notes, "other intergenerational mechanisms may operate to transmit values and behaviors from parent to

---

427. *Id.* at 175.

428. HARCOURT, *supra* note 323, at 57.

429. Robert E. Park, *The City: Suggestions for the Investigation of Human Behavior in the Urban Environment* (1925), in ROBERT E. PARK & ERNEST W. BURGESS, *THE CITY* 1, 45 (Heritage of Socio. ed. 2019).

430. *Id.*

431. Clark Tibbitts, *Success or Failure on Parole Can Be Predicted: A Study of the Records of 3,000 Youths Paroled from the Illinois State Reformatory*, 22 J. CRIM. L. & CRIMINOLOGY 11, 16 (1931).

432. Oleson, *supra* note 20, at 1365.

433. *Id.*

child.”<sup>434</sup> “Genetic influences, for example,” the guide continues, “may operate to transmit anti-social personality disorder and criminality.”<sup>435</sup>

A topic as fraught as the relationship between biology and crime requires more careful attention than a throwaway line tantamount to “just asking questions.” The idea that “criminality” is a discrete, genetically transmissible trait should be firmly situated in the dustbin of history. The substantial empirical literature on the relationship between genetics and crime points only toward a very modest connection with many caveats: The modest connection exists for certain types of offenders and certain types of crimes in the presence (or absence) of certain environmental factors.<sup>436</sup> Genetic predisposition plays a role in some subset of criminal activity,<sup>437</sup> but “criminality” cannot and should not be characterized as a genetic inheritance. Criminal risk assessment instruments and literature should make a clean break from eugenic theories.

D. Garbage-In-Garbage-Out-Garbage-In-Garbage-Out (Peer Association Measurement Problems)

The criminal associates risk assessment domain is fraught with issues of incentives, measurement error, magnitude, particularity, race, and social inequality. On the incentives front, instruments that rely on self-reports are confounded by the subject’s incentive to deny criminal associates,<sup>438</sup> but instruments that rely on assessor discretion (through interviews or record review) are confounded by measurement problems. As J.C. Oleson has explained, criminal records may tell an inaccurate story reflecting incomplete information, individual biases, systemic biases (such as higher arrest rates in communities more heavily targeted by law enforcement), and recording errors.<sup>439</sup>

Even if an assessment could produce an accurate count of criminal associates, the nature of the crimes and relationships matters.<sup>440</sup> There is a

---

434. COMPAS Guide, *supra* note 4, at 41.

435. *Id.*

436. See BONTA & ANDREWS, *supra* note 47, at 59-87 (providing an evenhanded overview of past and current research on potential biological bases of criminal behavior).

437. *See id.*

438. *See* Oleson, *supra* note 20, at 1369 (“[E]ven if the defendant somehow *does* know exactly how many of his friends are criminal peers, he is unlikely to reveal this information . . . . [I]t is simply not in the defendant’s interest to provide this information to the court.”).

439. *Id.*

440. Courts have raised this point in intimate association decisions. *See, e.g.,* Sponick v. Detroit Police Dep’t, 211 N.W.2d 674, 681-82 (Mich. Ct. App. 1973) (“The regulation does not proscribe only association with individuals recently convicted or currently suspected of some crime. The regulation prohibits a police officer from associating with a neighbor, fellow church members, etc., arrested once decades ago.”).

world of difference between, for instance, maintaining community with peers who smoke marijuana and maintaining community with peers who carry out armed home invasions. There is also a world of difference between an acquaintance who committed a crime twenty years ago and one who is actively committing crimes in the present day. Additionally, subjects who have “criminal acquaintances” for value-neutral or even laudable reasons—like doing anti-gang work, caring for family members with arrest or conviction histories, living in a halfway house, and working in businesses that hire former felons—are assessed no differently by some instruments than those who spend time with “criminal acquaintances” for the purpose of committing crimes.<sup>441</sup> And finally, the racial proxy issue rears its ugly head here, again. Black and Latinx communities are monitored, arrested, and convicted at higher rates than other communities.<sup>442</sup> The racial composition of defendants’ families and neighborhoods is, therefore, likely to bear on the proportion of their social contacts who have criminal histories.

The relationship between “criminal associates” and recidivism is widely accepted,<sup>443</sup> but researchers and instrument developers need to be far more critical of what *kinds* of associates are connected to recidivism. Instruments oversimplify the complexity of human relationships. ORAS-CST and its derivatives rely on an assessor’s vague aggregate estimate;<sup>444</sup> the LSI-R on an assessor’s report of the subject’s counts of the right and wrong kinds of associates;<sup>445</sup> COMPAS on the self-reported proportion of the subject’s associates who have past criminal justice contacts, gang affiliations, or substance use;<sup>446</sup> and the OST (which presents a simple binary of “Positive, law-abiding” versus “Negative, criminally oriented” peers) on the subject’s self-report or assessor’s estimation of whether their peers are, on balance, upstanding law-abiding boy scouts or low-down dirty dog criminal degenerates.<sup>447</sup> None of these approaches account for the “criminal associates” measurement concerns detailed above.

---

441. See *supra* Part IV.B.2.

442. See *Report to the United Nations on Racial Disparities in the U.S. Criminal Justice System*, SENT’G PROJECT (Apr. 19, 2018), <https://perma.cc/5NM3-53FE>.

443. See, e.g., Oleson, *supra* note 20, at 1353 (“The notion that criminal companions . . . might lead to criminal behavior . . . has found considerable support in empirical research.”).

444. COLLINS ET AL., *supra* note 71, app. E at 60 (ORAS-CST); ELEK ET AL., *supra* note 25, app. at A-79 (IRAS-CST); Texas Felony Score Sheet, *supra* note 354, at 3; Texas Misdemeanor Score Sheet, *supra* note 354, at 2.

445. Idaho LSI-R Scoring Guide—Version 3.0, *supra* note 57, at 16-17.

446. COMPAS Questionnaire, *supra* note 33, at 3.

447. STINSON, *supra* note 81, at 88 (OST self-report, which includes the “Positive, law-abiding” and “Negative, criminally oriented” language); *id.* at 99 (OST scoresheet, which presents a functionally identical choice between “Positive/prosocial” and “Negative (Criminal influence)” peer relationships).

Since there is no common agreement among instruments on how to measure peer associations, the instruments cannot be said to measure the same thing. In fact, there is empirical evidence that COMPAS and LSI-R do not: UCLA's 2010 COMPAS validation study found that COMPAS categorized 41% of subjects as having "High Problem Levels of Criminal Peers/Associates," while the LSI-R found that only 16% had high problem levels.<sup>448</sup> The authors further stated that "only 19% of those scoring high on the CASSPEER [COMPAS] scale also scored high on the Companions scale in the LSI-R."<sup>449</sup> They explained that COMPAS and LSI-R do not measure the same thing: "[T]he LSI-R scale represents a combination of social isolation and criminal peers, whereas the COMPAS scale focuses on the criminal/delinquent activity of the respondent's peers."<sup>450</sup>

Meta-analyses like the Gendreau study, performed on mismatched, oversimplified instruments, do not work because they create a twofold garbage-in-garbage-out problem. Instruments measure associates badly, then the meta-analysis aggregates the unreliable results of assessments that do not measure the same thing. The meta-analysis produces a number representing the supposed correlation between "criminal associates" and recidivism, but the number is meaningless because it relies on instruments employing differing methodologies, many of which are unsophisticated or poorly conceived.<sup>451</sup> The existence of a meta-analytic correlation does not mean any given instrument assesses the factor appropriately or that peer relationships can be reduced to a meaningful number for any given defendant.

Defendants' peer association scores are deceptively simple and, to keep playing a broken record, less useful than narrative explanation. Their empirical underpinnings are tenuous, they do not improve predictive accuracy,<sup>452</sup> and they obscure the complexity of the subject's relationships. Reliable assessment of any subject's associates requires accurate self-reports, as well as consideration of whether arrests without conviction count as criminal

---

448. FARABEE ET AL., *supra* note 153, at 16.

449. *Id.*

450. *Id.*

451. The looseness with which risk assessment developers treat this factor is also evident in the development of a risk assessment instrument not discussed in this Article, the Minnesota Screening Tool Assessing Recidivism Risk (MnSTARR). MnSTARR's developer wrote of the underlying research: "Reflecting the salience of antisocial associates as a recidivism risk factor, gang affiliation was predictive, at least for males, for several types of recidivism. Conversely, the effects of [prison] visitation for both males and females suggest that prosocial associates are a protective factor against reoffending." Duwe, *supra* note 136, at 597. This formulation fails to define "prosocial associates" and seems to assume, for no reason, that only "prosocial associates" visited the study's subjects in prison.

452. *See supra* Part II.C.1.



history, what kinds of crimes the associates committed, when they committed those crimes, and why the defendant associates with them. There is no reason to try to quantify that information.

## V. Beware the Shiny Legal Penny

In a 2018 concurrence, Iowa Supreme Court Justice Brent Appel cautioned against jumping on the risk assessment bandwagon too quickly:

Certainly the shiny legal penny of a new risk assessment tool should be carefully scrutinized by the courts. We should not forget the recent unattractive history in which the United States Department of Justice and the FBI, for over a decade, developed and advocated the use of “bullet match” analysis that was often presented by expert witnesses without providing the full picture of how statistically insignificant the “match” of the bullets really was. The relentless and potentially corrosive drive for efficiency and certainty in a resource-scarce public sector should not drive courts to use risk assessments in an unjustified “off label” manner or in a fashion that otherwise lacks meaningful empirical support to drive sentencing.<sup>453</sup>

Justice Appel’s caution is warranted. We need to beware of trends. The laboratories of democracy are churning out instruments at a steady clip, but criminal risk assessment could turn out to be no more than a passing fad, like the 1980s-90s mania for prison boot camps that faded over the past two decades, as one study after another revealed that they did not reduce recidivism.<sup>454</sup> Research on risk instruments’ predictive utility has been hampered by groupthink and the focus on individual predictions has overshadowed an application to which instruments are better suited: identifying systemic shortcomings and biases.<sup>455</sup>

---

453. *State v. Guise*, 921 N.W.2d 26, 32 (Iowa 2018) (Appel, J., concurring) (citations omitted).

454. See, e.g., Doris Layton MacKenzie, David B. Wilson & Suzanne B. Kider, *Effects of Correctional Boot Camps on Offending*, 578 ANNALS AM. ACAD. POL. & SOC. SCIS. 126, 137 (2001) (“In our overall meta-analysis of recidivism, we found no differences between the boot camp and comparison samples.”).

455. See Kristen Bell, Jenny Hong, Nick McKeown & Catalin Voss, *The Recon Approach: A New Direction for Machine Learning in Criminal Law*, 36 BERKELEY TECH. L.J. 821, 823 (2021) (laying out a roadmap for the “Recon Approach” to machine learning, which employs technological tools “to scrutinize how judges, parole board members, and other decisionmakers exercise discretion in the context of criminal law”); see also Josh Salman, Emily Le Coz & Elizabeth Johnson, *Florida’s Broken Sentencing System*, SARASOTA HERALD-TRIB. (Dec. 12, 2016), <https://perma.cc/B337-XVFA> (analyzing court and correctional data, including data on sentencing points, to identify racial disparities in Florida sentencing practices); Elizabeth Evans, *Applying Risk Management Principles to Court Oversight*, in TRENDS IN STATE COURTS 2015, at 29, 29-30 (Carol R. Flango, Deborah W. Smith, Charles F. Campbell, Elizabeth Maddox & Neal B. Kauder eds., 2015) (recounting how the Arizona Administrative Office of the Court constructed a risk assessment tool to identify court operations breakdowns and malfeasance with the assistance of the department that developed the OST).

COMPAS seems to be critics' favorite target, but it is less problematic than many other poorly developed and validated instruments currently flying under the radar. The OST is one such instrument: The developers knowingly included risk factors that are only weakly correlated with recidivism and assigned them equal weight to strongly correlated factors.<sup>456</sup> Washington, D.C. used an instrument for juvenile sentencing for fourteen years before defense attorneys "dug into the validity behind the system," and "found only two studies of its efficacy, neither of which made the case for the system's validity; one was 20 years old and the other was an unreviewed, unpublished Master's thesis."<sup>457</sup> Colorado's pretrial instrument is also an instructive example of risk assessment excess and mania. It was developed using a .30 threshold for statistical significance because, in the developers' words, "the sample size was too small to yield a sufficient number of predictors" to employ a standard .05 threshold.<sup>458</sup> That approach would be sloppy in an academic context, but it is totally irresponsible in a real-world context. Twenty-two Colorado counties now use the instrument for pretrial risk assessment.<sup>459</sup> Those counties evidently decided they were comfortable with a thirty percent probability that the instrument's predictions are meaningless. That is a careless way to play with people's lives. Risk assessment instruments should only be used if they are used wisely.

## Conclusion

Risk assessment instruments that include attitude and associational factors offer a false promise that quantification improves on narrative. Quantification causes more problems than it solves. If risk assessment instruments continue to include attitude and associational factors, stakeholders need to radically reconsider the instrument design process. They must examine who designs these things in the first place, how factors are assessed, and what factors are assessed.

*Who designs these things in the first place.* The lead developers of every instrument discussed in this Article appear to share a common race, gender, and level of education (white, male, doctorate), with only one exception—one of COMPAS's developers does not appear to have a PhD.<sup>460</sup> No developers

---

456. See *supra* Part I.D.2.

457. AI NOW INST., LITIGATING ALGORITHMS: CHALLENGING GOVERNMENT USE OF ALGORITHMIC DECISION SYSTEMS 14 (2018), <https://perma.cc/H3AT-VNQC>.

458. Stanford L. Sch. Pol'y Lab, Risk Assessment Factsheet: Colorado Pretrial Assessment Tool (CPAT) 2 (2019), <https://perma.cc/V99Q-JT3Q>.

459. *Id.* at 3.

460. See Tim Brennan, RESEARCHGATE, <https://perma.cc/2V6U-ZNHE> (archived Apr. 19, 2023) (to locate, select "View the live page"); Angwin et al., *supra* note 29 ("[Northpointe was founded] in 1989 by Tim Brennan, then a professor of statistics at the University of  
*footnote continued on next page*

disclose having consulted with people who have lived experience as defendants. As things stand, the developer pool does not represent the entire population and seems content to study criminal justice-involved people from a distance instead of soliciting their expertise.<sup>461</sup> The result is a hodgepodge of approaches to assessing attitudes and associations, none of which produce especially robust or defensible outcomes.

*How factors are assessed.* Scholars, policymakers, and practitioners must not only seek to understand the legal and empirical problems with existing risk assessment instruments, but also carefully scrutinize the content of risk assessment scoresheets and questionnaires for any instrument that comes down the pike. Although there are substantial barriers to transparency in criminal risk assessment, a lot of information is available if someone digs for it. The sub-surface problems with risk assessment should not be overlooked just because the surface problems are pressing and easier to observe.

*What factors are assessed.* Attitude and associational factors stand on shaky empirical, normative, and legal ground. They do not improve predictive validity—in fact, they appear to lessen it. They are neither highly correlated with recidivism nor necessarily causal. And they do not assist with needs assessment, since narrative explanations are more useful and precise than need scores. On the legal front, items that relate to abstract beliefs or expressive associations are likely to infringe on First Amendment rights at sentencing and “criminal family” items are likely to violate Fourteenth Amendment equal protection. Peer association items are less likely to infringe on constitutional rights, but they still have little predictive value and are bedeviled by measurement problems. There could be a universe in which attitude and associational factors achieve more predictive power and remain within constitutional bounds, but we are not there yet. Until we get there, the factors should have no place in risk assessment.

---

Colorado, and Dave Wells, who was running a corrections program in Traverse City, Michigan.”); Donald Andrews, OTTAWA CITIZEN (Oct. 26, 2010), <https://perma.cc/Y5WJ-AL2Y> (“Don completed his doctorate degree in psychology at Queen’s University in 1969, after which time he joined the faculty at Carleton University, where he remained throughout his academic career.”); James Bonta, LINKEDIN, <https://perma.cc/59SW-MJKF> (archived Apr. 19, 2023) (to locate, select “View the live page”) (“Jim Bonta received his Ph.D. in Clinical Psychology from the University of Ottawa in 1979.”); David Simourd, CARLETON UNIV., <https://perma.cc/756Y-26B8> (archived Apr. 19, 2023) (“Dr. David Simourd graduated from Carleton with a Ph.D. in Psychology in 1992.”); Nicole Washington, *The School of Criminal Justice Remembers Longtime Correctional Policy Expert*, UNIV. CINCINNATI (Mar. 8, 2022), <https://perma.cc/9HFX-H7S7> (to locate, select “View the live page”) (“The faculty and staff in the School of Criminal Justice . . . mourns the passing of its ‘larger than life’ former school director, Edward Latessa, PhD.”).

461. See Ngozi Okidegbe, *The Democratizing Potential of Algorithms?*, 53 CONN. L. REV. 739, 774-80 (2022) (advocating for shifting power to communities most impacted by criminal justice policy).