COMPAS-R Core

The Revised Version of the Standard COMPAS Core

Technical Manual
For COMPAS-R Core

Research Department

Northpointe, Inc.

March 28, 2022

Contents

1 Introduction

This technical manual for the COMPAS-R Core contains information on the COMPAS-R scale refinement process and the methods used to score, interpret and evaluate the COMPAS-R scales. The manual describes the COMPAS-R norm groups and the procedures used to assign cases into Low, Medium and High (or Unlikely, Probable, and Highly Probable) classifications. It also presents analyses that examine the distributions and other properties of the scales in the COMPAS-R norm groups, such as scale means and standard deviations. In addition, the manual provides results from analyses that demonstrate the internal consistency, reliability, and predictive validity of the COMPAS-R scales.

This work and the resulting scales build on the original work performed during the development, maintenance, and refinement of the Standard COMPAS Core that continues to be available in the Northpointe Suite Software. The goals for the COMPAS-R were to shorten the Standard COMPAS Core scales, remove scales that were not critical to understanding a person's criminogenic needs, modify the calculation of the General Recidivism Risk scale so that it could be done by hand, and make items and scale scoring transparent and available for public inspection.

2 Description of the COMPAS-R Core Norm Groups

Beginning with COMPAS version 5.0, and continuing with COMPAS-R, the default method for classifying scale scores as Low, Medium, or High (or Unlikely, Probable, and Highly Probable) is based on a gender-specific Composite norm group. The agency also has the option of selecting an appropriate gender-specific subgroup within the composite norm as a reference group when configuring their scale sets. For example, in a probation setting, the user could compare a Woman's COMPAS scale score to the Women's Composite norm group or the Women's Probation subgroup within the composite norm group (Brennan, Dieterich, and Oliver 2007). Details about the composition of the Composite norm group and subgroups are provided in Appendix 10.1.

3 Overview of COMPAS and COMPAS-R

The structure and much of the content of this report is based on earlier COMPAS reports such as the *COMPAS Technical Manual and Psychometrics Report Version 5.01*, Brennan, Dieterich, and Oliver (2007). From that report:

> Statistically based risk/needs assessments have become accepted as established and valid methods for organizing much of the critical information relevant for managing persons

> in correctional settings (Quinsey et al. 1998). Many research studies have concluded that objective statistical assessments are, in fact, superior to human expert judgment (Grove et al. 2000; Swets, Dawes, and Monahan 2000). COMPAS is a statistically based risk assessment specifically developed to assess many of the key risk and needs factors in adult correctional populations and to provide decision-support information regarding the placement of persons in the community. It aims to achieve these goals by providing valid measurement and succinct organization of many of the salient and relevant risk/need dimensions. A further goal is to help practitioners design case-management support systems for persons in community placements.

COMPAS-R Core offers an alternative to COMPAS Core, building on its effectiveness, brevity, and ease of administration, while being shorter and more transparent. Additionally, the administration of the COMPAS-R Core within the Northpointe Suite software has been completely revised, offering agencies and users alternative organization options and reports that make scoring more transparent and easier to understand. The standard COMPAS Core and software interface remain available to users who prefer it.

Any items that were not retained from the standard COMPAS Core for the COMPAS-R are available in the software within "Non-scored Domain items." These may be selected for inclusion in assessments at the time of scale set configuration. COMPAS-R Core scales and domains are listed here:

- **Criminal Involvement Domain**
  - Legal System Involvement scale
  - History of Non-compliance scale
  - History of Violence scale
  - Current Violence scale
  - Non-scored Criminal Involvement Domain items
- **Relationships/Lifestyle Domain**
  - Associates and Peers scale
  - Substance Use scale
  - Drug Problems scale
  - Society and Routines scale
  - Leisure and Recreation scale
  - Non-scored Relationships/Lifestyle Domain items
- **Personality/Attitudes Domain**
  - Thinking and Attitudes scale
  - Personality Traits scale
  - Anger scale
  - Cognitive Behavioral scale
  - Non-scored Personality/Attitudes Domain items
- **Family Domain**
  - Socialization History scale
  - Non-scored Family Domain items
- **Social Exclusion Domain**
  - Financial scale
  - Vocational/Educational scale
  - Residential Stability scale
  - Non-scored Social Exclusion Domain items

Validity Tests in COMPAS-R Core Concerns over the quality and integrity of criminal justice data and the reliability of respondents in interview settings motivated the creation of built-in data validity tests for the standard COMPAS Core. These concerns are addressed in the COMPAS-R Core by utilizing a test for random responses that looks for possible inconsistencies. Once a certain percent of potential inconsistencies are identified, a warning is provided during the score calculation phase in the software.

Additionally, there are internal validation checks built into the software to assist the practitioner in generating consistent and accurate responses in the Official Records section during data input. A warning would be issued, for example, if the age at first arrest entered is greater than the person's current age.

4 Methods Overview

Two methods were used to shorten the COMPAS need scales. These methods identified items in the data that did not contribute substantially to the scale's performance, based on their respective methodological criteria. Some of the identified items that preserve critical information were retained.

The first method examines whether the need scales meet the assumptions of Classical Test Theory (CTT). Scales that meet the assumptions required for CTT analysis underwent additional analysis to determine which, if any, items can be removed and still be acceptable measures of the latent construct, that is, the need to be identified. Many of the recommendations from Wieland et al. (2017) were followed. These largely echo the refinement process that the scales went through in their original development. At that time, they were evaluated in terms of Theoretical-Substantive Validation, Factorial Validation, and External Criterion-Related Validation.

From Brennan, Dieterich, and Oliver (2007):

> Each scale was developed as a unidimensional linear function and should thus conform to the psychometric requirements of such scales. In this phase, standard factor analytic and item analysis techniques were used to examine the contribution of each item to the factorial validity of each scale. Items were dropped if they did not load or correlate with the intended scale. The items that survived this phase were then included in subsequent stages of validation. This phase established both the internal reliability of each scale (Cronbach's alpha) and the factorial validity of the scales.
>
> In general, items were selected to be both theoretically related to a construct and to load heavily on the same component or factor. The factor pattern for each scale is presented in Chapter 7 – see, for example, Table 5. These factor patterns represent the rotated principal component analyses for each scale.
>
> The importance of each scale item is indicated by the size of the factor loading: A large loading by an item on the first component (e.g., 0.6 and higher) indicates that the item is an important contributor to the scale. By including several such items with strong loadings on each scale, we will achieve a large eigenvalue for the first factor, or component. A large eigenvalue (greater than 1) indicates that the factor accounts for a large amount of overall variance in the summated score, which, in turn, suggests that the scale is measuring a single underlying construct. It is important to note, that the loadings derived from principal component analysis will tend to be biased upwards (Fabrigar et al. 1999), so it is important to focus on the relative magnitudes of the

loadings.

Summative General Recidivism Risk Scale (Summative GRRS) Modifying the GRRS involved re-fitting the model to new data using inputs from the newly revised need scales. After the new model was developed, it was transformed to a summative scale, following the ideas presented in Sullivan, Massaro, and D'Agostino Sr. (2004). For details on the development of this scale, see Chapter 7.16.

4.1 Item Response Theory (IRT)

The previous section discussed *Classical Test Theory* (CTT), commonly used for assessing psychological tests. IRT can be used in conjuncture with CTT. IRT analysis for this report included assessing the COMPAS Core scales and keeping items such that scales would follow the non-parametric monotone homogeneity model.

The purpose of checking if an IRT model describes the data well is to use the measurement properties implied by a particular IRT model. The *ordinal person measurement* property is one of particular interest since scale scores are computed by adding items (items from a questionnaire: CORE-R Need/Risk Scales) together and interpreting the score as some type of measure. The *ordinal person measurement* property implies that scores are distinguishable and have meaning: this is called stochastic ordering of θ (the latent trait being measured) by $X+$ (the test/scale score when all items are added up). For dichotomous items, this means,

$$P(\theta > a|X_+ = L) \geq P(\theta > a|X_+ = K) \forall a \text{ and } \forall K < L.$$

(Hemker et al. 1996; Grayson 1988; Huynh 1994). The non-parametric monotone homogeneity model implies the *ordinal person measurement* property (Van der Ark and others 2012), for dichotomous items. For polytomous items, the monotone homogeneity model implies weak stochastic ordering:

$$P(\theta > a|X_+ > K) \geq P(\theta > a|X_+ \leq K) \text{ for all } a \text{ and for all} K.$$

The Monotone Homogeneity Model implies the following with respect to the data:

1. Unidimensionality: Only one latent variable θ is required to explain the association between item scores.
2. Local independence:

$$P(X_1 = x_1, ..., X_J = x_J|\theta) = \prod_{j=1}^{J} P(X_j = x_j|\theta)$$

3. Latent monotonicity

$$P(X_j \geq x|\theta_a) \leq P(X_j \geq x|\theta_b), \text{ for all } \theta_a < \theta_b \text{ for } j = 1, ..., J;\ x = 1, ..., m$$

where X_j denotes the score for item j with realizations $x_j = 0,1, ..., m$.

Unidimensionality is tested using Principal Components Analysis (PCA). We test the assumption of local independence by performing a chi-squared test of independence on all item pairs for each of

the COMPAS-R Core scales. The latent monotonicity assumption is tested using the `R` function `check.monotonicity` from the package in **mokken** (Van der Ark and others 2007).

The function summary(`check.monotonicity(data)`) returns a matrix, displaying for each item a summary of the results of and checks of manifest monotonicity (Van der Ark and others 2007). The first column lists all the items. The second column, `ItemH` contains the H_j values for each item. The column `#vi` displays the number of item violations; `#vi/#ac` displays the percent of violations for all comparisons made for that item. The column `#zsig` reflects whether the violation is significant.

Most of the IRT analysis was also performed on the original COMPAS Standard scales and used as comparison to the final results on the items selected for each analogous scale in COMPAS-R Core.

4.1.1 Mokken Scale Analysis (MSA)

Mokken Scale Analysis (MSA) is used to investigate the fit of non-parametric IRT models (Van der Ark and others 2012). We asses three types of scalability coefficients that play an important role in MSA. Checking the MSA assumptions provides insight into what items can be kept that will help the data to follow the MHM. If the MSA assumptions are not met, the MHM assumptions will also not be met. The three scalability coefficients that are used for the data are discussed below. All information can be found in Van der Ark and others (2012).

Item-pair scalability coefficient The item-pair scalability coefficient, H_{ij}, is defined as the normed covariance between item scores when other assumptions are met. The Monotone Homogeneity Model (MHM) implies that the item-pair scalability coefficient is positive, with higher values indicating a stronger scale. Pairs of items with a negative item-pair scalability indicate that at least one of the items does not fit the MHM and should be removed.

Item scalability coefficient The item scalability coefficient, H_j is defined as the normed covariance between the item score and the rest score when other assumptions are also met. The rest score is the total score minus X_j. If the data follows a MHM, then $0 \leq H_j \leq 1$. For our analysis, we use a lower bound of $c = 0.30$ for all item scalability coefficients.

Test-scalability coefficient If the data follow the MHM, the test-scalability coefficient, H, is between zero and one. If $0.3 \leq H < 0.4$, then we say the scale is weak. If $0.4 \leq H < 0.5$, then we say the scale is moderate, and if $H > 0.5$, the scale is considered strong.

Discussion Developing Mokken scales in not an easy task. Most of the data did not follow the MHM for any of the scales. However, in our analyses, we try to model each scale such that the data would follow the Monotone Homogeneity model: the MHM has many attractive qualities such as the ordinal measurement property briefly explained above.

Although many of the scales were able to be modeled such that they met the MSA assumptions, most failed the test of item independence when the IRT assumptions were tested. There were four scales with at least one pair of items that were found to be independent. The four scales are:

1. History of Violence
2. Current Violence
3. Socialization Failure

4. Vocational/Educational

The Current Violence scale exhibited the most number of pair of items that demonstrated independence with the chi-squared test of independence in `R`. However, it did not meet the latent monotonicity assumption and failed the MSA assumptions: negative item-pair scalability coefficients were present; many of the item-scale coefficients were less than the lower bound of $c = 0.03$; and the test scale coefficient (0.245) indicated the scale was very weak. The History of Violence scale also had many item scalability coefficients that were too low; only two items had an item scalability coefficients greater than or equal to $c = 0.03$ and the test scalability coefficient (0.263) indicated a weak scale. The Socialization History scale only had two items with item scalability coefficients greater than or equal to 0.03 as well; the test scalability coefficient was $H = 0.266$ implying a very weak scale. The Vocational/Educational scale is considered a higher order scale; all MSA and MHM IRT analysis that was completed on most of the other scales, was still performed on this scale, but it is not surprising that many of the assumptions were not met.

The Legal System Involvement scale did not demonstrate independence for any of its items; however, the MSA statistics provided some very encouraging results with respect to the MHM. All scalability coefficients were above 0.50 and the smallest item scalability coefficient was 0.570 (***n_prison***). Furthermore, the test scalability factor was $H = 0.601$, implying a very strong scale. The PCA results demonstrated that the scale was unidimensional and there were no item violations when the latent monotonicity assumption was tested.

The Residential Stability and Leisure & Recreational scales also parred well in the mokken scale analysis. The Residential Stability scale had a range of $0.234 - 0.858$ for the item-pair scalability coefficients. The range for the Leisure & Recreational was $0.654 - 0.855$. The smallest item scalability coefficient for Residential Stability was 0.313 and the largest was 0.658 for the item ***yrs_address***; the test scalability coefficient was 0.528. The item scalability coefficients for the Leisure & Recreational scale were exceptionally high; all values were between 0.693 and 0.754 with $H = 0.733$.

5 Scoring and Interpreting the Scales

5.1 Basic Descriptive Information for the Scales

Two types of scales comprise the set of COMPAS-R scales:

1. Basic scales that measure need constructs (e.g., associates, residential stability) or basic risk factors (e.g., criminal history). These are not meant to be predictive but aim simply and accurately to describe the individual in terms of relevant needs and history.
2. A general recidivism risk scale that integrates the basic risk scales to provide a predictive measure of a person's risk of recidivism at the time of assessment. General recidivism is defined for this scale as a new arrest within three years following a COMPAS-R assessment administered at the time of community start.

Table 1 lists all scales in the COMPAS-R. Each scale is described in Chapter 7 of this report. The items entering into each scale are also described in that section.

The number of observations (N), means, standard deviations, and alpha coefficients for the scales are also shown in Table 1. Note that N varies from one scale to the next because of missing values or responses that have been excluded in the new scale. Scales with missing item values are not scored, so for a given scale, if any of the responses were missing or excluded, the value for the scale was considered missing.

Table 1: Summary of COMPAS-R Core scales in the COMPAS-R normative data.

	Items	N	Min	Max	Mean	SD	Alpha
Summative GRRS	19	12426	11	56	36.08	7.13	a
Legal System Involvement	4	12426	0	15	5.87	3.59	0.71
History of Non-compliance	5	12426	0	15	1.87	2.37	0.73
History of Violence	6	15731	0	13	1.79	2.22	0.63
Current Violence	7	16011	0	6	0.85	1.12	0.52
Associates and Peers	5	16089	0	13	2.5	2.69	0.77
Cognitive Behavioral	35	4314	0	76	20.08	9.85	0.83
Society and Routines	12	11851	0	21	5.81	3.69	0.62
Leisure and Recreation	4	12761	0	6	1.46	1.74	0.81
Thinking and Attitudes	9	9682	0	26	7.09	4	0.9
Personality Traits	10	8766	0	27	10.19	4.24	0.81
Anger	5	8766	0	15	4.53	2.5	0.78
Socialization History	10	7764	0	21	4.75	3.55	0.68
Financial	4	11204	0	8	2.86	2.06	0.73
Vocational/Educational	8	21114	0	13	5.24	3.16	0.67
Residential Stability	6	14568	0	14	4.27	3.19	0.64
Substance Use	5	17926	0	8	2.29	2.16	0.71

[a] Alpha not available because this scale is based on a linear equation.

All scales are summative. Details regarding their point values are contained in the corresponding sections of Chapter 7. Note that some scales may not have any members of the norm set that obtain

the minimum or maximum possible value. These scales are Current Violence (maximum possible points: 7), Thinking and Attitudes (maximum possible points: 27), Personality Traits (maximum possible points: 30), Cognitive Behavioral (maximum possible points: 80), Socialization History (maximum possible points: 22), and Summative GRRS (0 to 71 possible points).

5.2 Conversion of Raw Scale Scores to Text Levels

The text levels of the COMPAS-R Core may be either "Unlikely," "Probable," and "Highly Probable," for the need scales, or "Low," "Medium," and "High," for the scales that measure criminal history and the Summative GRRS. In the COMPAS-R Core, scale scores can be referenced to the scale distributions of eight normative groups:

1. Male Composite,
2. Male Prison (including Parole),
3. Male Probation,
4. Male Jail (client-specific),
5. Female Composite,
6. Female Prison (including Parole)
7. Female Probation, and
8. Female Jail (client-specific).

In Chapter 7 of this report, we also provide score ranges for "Gender Neutral" normative groups in the four placement groups, but those are not available in the Northpointe Suite at this time. They are composed of the corresponding Male and Female norm groups.

The text levels indicate whether a score is high, medium, or low *with respect to* other members of the norm group. Many of the scales that are associated with official records/criminal history (including the Summative GRRS) are divided into the lowest 40% of scores, the next 30% of scores, and the highest 30% of scores. The need scales Associates and Peers and Cognitive Behavioral, follow this scheme as well. However, most of the need scales are divided into the lowest 50% of scores, the next 20% of scores, and the highest 30% of scores.

The cutting scheme used for the Substance Use scale is unlike the other scales in the COMPAS-R Core in that only the lowest 20% of scores are considered low and the next 20% are considered medium. This is because the scale is designed to *screen* for the possible presence of a substance use problem. A person who scores in the Probable range would be recommended for further evaluation (e.g. UNCOPE, ASI, etc.) and a person who scores in the Highly Probable range may have a serious alcohol or drug problem requiring a structured treatment approach. Because of the high incidence of drug/alcohol abuse within the criminal justice population, assisting the person to attain and maintain sobriety is a primary intervention to impact recidivism. For more information, see section 4.2 of the Practitioner's Guide to COMPAS-R Core (Northpointe Inc. 2021).

Table 2: COMPAS-R Core scales, text levels, and corresponding Percentages in the normative data.

Scale	Text Levels	Score Range Percentages
Summative GRRS	Low, Medium, High	40–30–30
Legal System Involvement	Low, Medium, High	40–30–30
History of Non-compliance	Low, Medium, High	40–30–30
History of Violence	Low, Medium, High	40–30–30
Current Violence	Low, Medium, High	40–30–30
Associates and Peers	Unlikely, Probable, Highly Probable	40–30–30
Cognitive Behavioral	Unlikely, Probable, Highly Probable	40–30–30
Society and Routines	Unlikely, Probable, Highly Probable	50–20–30
Leisure and Recreation	Unlikely, Probable, Highly Probable	50–20–30
Thinking and Attitudes	Unlikely, Probable, Highly Probable	50–20–30
Personality Traits	Unlikely, Probable, Highly Probable	50–20–30
Anger	Unlikely, Probable, Highly Probable	50–20–30
Socialization History	Unlikely, Probable, Highly Probable	50–20–30
Financial	Unlikely, Probable, Highly Probable	50–20–30
Vocational/Educational	Unlikely, Probable, Highly Probable	50–20–30
Residential Stability	Unlikely, Probable, Highly Probable	50–20–30
Substance Use	Unlikely, Probable, Highly Probable	20–20–60

5.3 Interpreting Text Levels

Text levels can only be interpreted relative to the norm sample from which they were developed. If the sample on which, say, a recidivism risk scale is based happens to consist exclusively of individuals with high propensity to recidivate, then low text levels do not necessarily indicate a low risk of recidivism. Similarly, if the sample happens to consist mainly of individuals with low propensity to recidivate, the text levels for the risk scale would be biased in the other direction – high scores could be associated with individuals who are actually not very likely to commit new crimes.

Another important caveat is that for some scales, it is not always possible to break the sample into three groups of the desired ratios, necessitating that the middle text level be skipped over. For example, in the Female Probation norm group, only Low and High were possible for the Current Violence scale (see Table 45). This is due to the fact that a very large proportion of women in the Composite norm scored 0 on this scale, subsuming both the lowest 40% of scores and the middle 30% of scores. Only the top 30% of scores remained to make up the Highly Probable text level. This effect is can be seen in the cumulative distributions for the Current Violence scale displayed in Figure 7. The figure shows that over 70% of the cases in the Female Probation norm group had a score of 0 on this scale. The distributions of the Composite norm group and the Jail norm sub-group for women behave similarly. The other COMPAS-R scale scores whose distributions are affected this way for some norm groups or subgroups are Substance Use (Female Composite and Jail), Anger (Male Jail), History of Non-compliance (Conviction Version, Female Probation and Composite), and History of Violence (Conviction Version, all Female groups).

When it was not possible to divide the sample into the exact ratios desired, we identified ranges that divided the scores into groups as close as possible to these ratios, rounding to the lower text

level if a score straddled 2 ratio cuts. For example, suppose that a scale to be cut into ratios of 40-30-30 had 68% of the norm group scoring 10 or lower, while 75% scored 11 or lower. Then, a score of 11 would be considered a middle text level for that scale, while scores of 12 and higher would be in the high text level.

5.4 Norm Groups and Scale Distributions

In COMPAS-R Core, the user has the option of selecting the gender-specific the composite norm or a gender-specific subgroup within the composite norm as a reference group. In other words, a person's scale score will be converted to a text level by comparing it with the scores of people of the same sex in the corresponding composite norm or in the selected subgroup within the composite norm. The composite norm group and all subgroups are described in Appendix 10.1.

As an example, in a probation setting the user could compare a Woman's particular COMPAS-R scale score to either the Female Composite norm group or to the Female Probation subgroup of the composite norm group. The COMPAS-R provides these optional norm groups to provide some control over the relative meaning of the scale text levels. Text levels are based on the distributions in a normative group and the relative meaning of a particular text level is dependent on the norm group that is referenced.

In Chapter 7, the distribution of the scale scores in each of the norm groups is examined. Here we provide a brief description of the graphs and plots of the scale distributions that are presented in that section.

Figures 1 through 30 show graphs and plots that compare the distributions of the COMPAS-R scale scores of males and females in the COMPAS-R norm groups. For each COMPAS-R scale, plots showing the cumulative distribution and histograms are presented. The cumulative distribution of the COMPAS-R scale score shows the percent of individuals at or below each score for the scale. For example, a Legal System Involvement scale score of 5 is located at the 60th percentile of the distribution of scores for females in the probation norm group, as shown in the lower right panel of Figure 1. For females in the prison/parole norm group, a score of 5 is located between the 20th and 40th percentile, as shown in the lower left panel of the same figure.

The histograms show the distribution of scores for each scale. The histograms for the Legal System Involvement scale shown in Figure 2 depict the score distributions for males and females in the COMPAS-R norm groups. The histograms are scaled to density units. This means that the heights of the bars are scaled so that the sum of the areas of all the bars is equal to 1. A normal density curve based on the mean and standard deviation of the scale scores is drawn over the histogram of each panel. The histograms in Figure 2 show that a higher proportion of females in the prison/parole norm group have scores of 10 or above on the Legal System Involvement scale, in comparison to females in the probation norm group. The overlay of the normal density curve shows that the scores for females in the probation norm group are positively skewed, with an excess of high scoring cases to the left and a relatively long and shallow tail of higher scoring cases to the right.

6 Evaluating the Scales

The validity and reliability of the scales were examined with methods that are widely used in the fields of measurement and test assessment. Reliability concerns how consistent a scale would be if the same subjects were repeatedly tested. Validity, on the other hand, can be measured in various ways. In this chapter, we focus on criterion-related validity and construct validity (DeVellis 2016).

6.1 Reliability

For a scale to be useful it must be reliable: if one were to carry out repeated testing of a given respondent with different questions or tests, approximately the same scale value should be obtained on each re-test. Generally, if the items entering a scale are sufficiently correlated (internally consistent), then the summated scale will be reliable. Cronbach's alpha provides a convenient measure of the reliability of a scale. Cronbach's alpha for each scale was calculated in the COMPAS-R composite norm group (all genders). The coefficients for the COMPAS-R composite norm group appear in Table 1. In the COMPAS-R composite norm group, 10 of 16 COMPAS-R scales have alpha coefficients of .70 or higher. By convention alphas of .70 and above indicate good internal consistency.

History of Violence and Current Violence are scales that inventory aspects of a person's criminal record, so the fact that their alpha coefficients are less than .70 is not of great concern. Alpha coefficients for the COMPAS-R scales Society and Routines, Socialization History, Residential Stability, and Vocational/Educational are slightly lower than 0.70, as well. These values, however, are either the same or marginally different than the alphas in the data before item removal or other revisions. For example, the value of alpha is partly a function of the number of items in a scale. When items are combined to shorten the scale, a slight reduction in alpha is expected. Two items in the Socialization History scale were combined into a single item, reducing alpha from .70 to .68. The Residential Stability scale had three items combined into one, reducing alpha from .68 to .64. Discussion about the change in alpha for each scale is contained in the corresponding section of Chapter 7.

In Chapter 7, tables for each scale are presented in which alphas are calculated when individual items are successively removed from the scales (see, for example, Table 7. Much like the tables with the scale factor patterns, these tables can identify items that make strong contributions to the scale. The Item-Total correlations also appear in these tables. These values represent the correlation between an item and a scale after the item has been subtracted from the scale. A high value indicates that the item strongly contributes to the reliability, or internal consistency, of the scale. Occasionally, items with less than ideal correlations are retained because of their interest for case management, such as the item that asks if the person is currently a gang member from the Associates and Peers scale. It has a correlation of 0.265 with that scale, but it is important information for case management. Removing this item from the scale would increase the alpha coefficient to 0.809, but from an already high reliability coefficient of 0.768. See Table 54.

6.2 Predictive Validity

The COMPAS-R Summative GRRS is used to assess risk of new arrest within three years of community start. Hence, it is important that the scale has good predictive validity. To assess the predictive validity of a scale, one approach is to measure a group of individuals on the scale, and then follow those individuals over time to determine whether the scale predicts, or is significantly associated with the behavior of interest. Details about the predictive validity of the Summative GRRS are provided in Chapter 7, Section 7.16.

7 Descriptions and Analyses of the Scales in the COMPAS-R Norm Groups

In this chapter, we give details about the analyses performed to derive the COMPAS-R Core scales from the standard COMPAS Core scales. Psychometric properties of the standard COMPAS Core scales (Brennan, Dieterich, and Oliver 2007) are reviewed and we demonstrate the degree to which those properties have been maintained in the COMPAS-R Core scales.

Some of the names of the standard COMPAS Core scales are changed in the COMPAS-R Core to reflect agency requirements. In general, name changes express greater neutrality, although the scales themselves continue to measure approximately the same need or risk construct. For example, in the standard COMPAS Core, the scale that measures one's degree of involvement with anti-social peers is called "Criminal Associates and Peers," whereas in the COMPAS-R Core, it is called "Associates and Peers."

7.1 Legal System Involvement Scale*

Both the standard and conviction versions of the Legal System Involvement scale are derived from the respective standard and conviction versions of the Criminal Involvement scale in the standard COMPAS Core.[1] In the development sample, there were 12,426 individuals with completed Criminal Involvement scales that were used to derive the standard version of the Legal System Involvement scale. 5,299 individuals had completed a conviction version Criminal Involvement scale.

The Legal System scale score is also an input for the Summative General Recidivism Risk Scale.

Review of Theoretical and Substantive Validity

As stated in the original psychometric studies of COMPAS, criminal involvement is a major risk factor in predicting recidivism. Various meta-analytical studies support this finding (Andrews and Bonta 2016; Gendreau, Goggin, and Little 1996). Juvenile delinquency has also been linked to future criminal behaviors (Moffitt 1993).

Scale Items

The standard and conviction versions of the Legal System Involvement scale are constructed using four items from the Criminal Involvement domain of the COMPAS-R Core. These items and their short descriptions are shown in Tables 3 and 4.

[1] Scales denoted with an asterisk (*) have both standard and conviction versions.

The Legal System Involvement scale items focus primarily on criminal activity engaged in during a person's adult life. The item asking for number of previous arrests (standard version) is the only item that includes events that occurred while the person was either an adult or a juvenile. Its counterpart in the conviction version of the scale asks about the number of previous convictions and adjudications that occurred while the person was either an adult or a juvenile.

The analyses that follow apply only to the standard version of the Legal System Involvement scale.

Table 3: COMPAS-R Core Legal System Involvement Scale Items (Standard Version). Practitioners are asked to verify all responses with the individual's criminal record. Any additional information that is shared by the individual that cannot be verified should be noted and scored. Exclude the current case for all questions.

Item	Short Description (Response Categories)
jail30_R	Has this person ever been sentenced to jail for 30 days or more as an adult (exclude current)? *(No=0; Yes=1)*
n_prison_R	How many times has this person been sentenced (new commitment) to a state or federal prison as an adult (exclude current)? *(0=0; 1=1; 2=2; 3=3; 4=4; 5+=5)*
n_probations_R	How many prior times has this person been sentenced to probation as an adult (exclude current)? *(0=0; 1=1; 2=2; 3=3; 4=4; 5+=5)*
t_prev_arrests_R	How many prior times has this person been arrested as an adult or juvenile (including for possession of small amounts of marijuana)? *(Text entry: 0 times=0; 1 time=1; 2–3 times=2; 4–5 times=3; 6+ times=4)*

Table 4: COMPAS-R Core Legal System Involvement Scale Items (Conviction Version). Practitioners are asked to verify all responses with the individual's criminal record. Any additional information that is shared by the individual that cannot be verified should be noted and scored. Exclude the current case for all questions.

Item	Short Description (Response Categories)
jail30_R	Has this person ever been sentenced to jail for 30 days or more as an adult (exclude current)? *(No=0; Yes=1)*
n_prison_R	How many times has this person been sentenced (new commitment) to a state or federal prison as an adult (exclude current)? *(0=0; 1=1; 2=2; 3=3; 4=4; 5+=5)*
n_probations_R	How many prior times has this person been sentenced to probation as an adult (exclude current)? *(0=0; 1=1; 2=2; 3=3; 4=4; 5+=5)*
t_prev_conv_R	How many prior times has this person been convicted as an adult or juvenile (including for possession of small amounts of marijuana)? *(Text entry: 0 times=0; 1 time=1; 2–3 times=2; 4–5 times=3; 6+ times=4)*

Item changes from the Criminal Involvement scale in the standard COMPAS Core

The item on the standard COMPAS Core that asks how many times a person was jailed for more than 30 days was changed to a dichotomous item asking whether the person had ever been sentenced to jail for more than 30 days. This change reduced the reliability of the scale from 0.782 to 0.714, but it is still within an acceptable range of reliability (Cortina 1993). Additional statistics for the scale are included in this section.

Factor structure in the norm data

Table 5: Principal component loadings.

	PC1	PC2	PC3	PC4
n_prison	-0.463	-0.579	0.669	0.054
n_probations	-0.698	-0.220	-0.679	0.052
t_prev_arrests	-0.527	0.782	0.300	0.144
jail30	-0.139	0.070	0.044	-0.987

Table 6: Eigenvalues and the proportion of the variance explained by each principal component.

	PC1	PC2	PC3	PC4
Eigenvalue	3.788	1.135	0.885	0.162
Standard deviation	1.946	1.065	0.941	0.403
Proportion of variance	0.634	0.190	0.148	0.027
Cumulative proportion	0.634	0.825	0.973	1.000

Principal components analysis of the Legal System Involvement items in the COMPAS-R composite norm group is presented here. The loadings of the principal components are shown in Table 5. Table 6 shows the eigenvalues of the principal components and the proportion of the variance explained by each. As shown in the table, the first principal component has an eigenvalue of 3.788 and explains 63.4% of the variance in the data. Three principal components are required to explain at least 90% of the variation in the data. This is comparable to number required in the standard COMPAS Core data.

Scale consistency and reliability in the norm data

Table 7: Correlation of items with scale score, and alpha reliability of the scale when each item is removed. Compare with alpha of 0.714 for all four items.

	Correlation with score	Alpha when item removed
n_prison	0.744	0.648
n_probations	0.847	0.575
t_prev_arrests	0.778	0.631
jail30	0.628	0.711

Cronbach's alpha measures a scale's internal consistency and reliability. Alpha for the Legal System Involvement scale in the composite norm set is 0.714. Table 7 shows the correlation of each item with the overall scale score, and the value of alpha when the item is removed from the scale. Removing the jail item affects alpha minimally, but it is an important item to have information about, and it is well-correlated with the scale score. Removing any of the other items would lower alpha more substantially.

Scale distributions in the norm groups

Table 8: Tabulation by population and gender with mean scores. Note: this table does not present the scale score from the single transgender person in the data.

Group	N	Mean
Jail Female	405	6.4
Prison Female	856	6.4
Probation Female	1561	4.0
Jail Male	1900	7.0
Prison Male	3350	6.7
Probation Male	4353	5.3
ALL	12425	5.9

Table 8 shows the mean Legal System Involvement scale score for each population subgroup in the normative data. The relationships among these means are as expected, with Women generally lower when comparing scores by gender, and people on probation generally lower when comparing correction types.

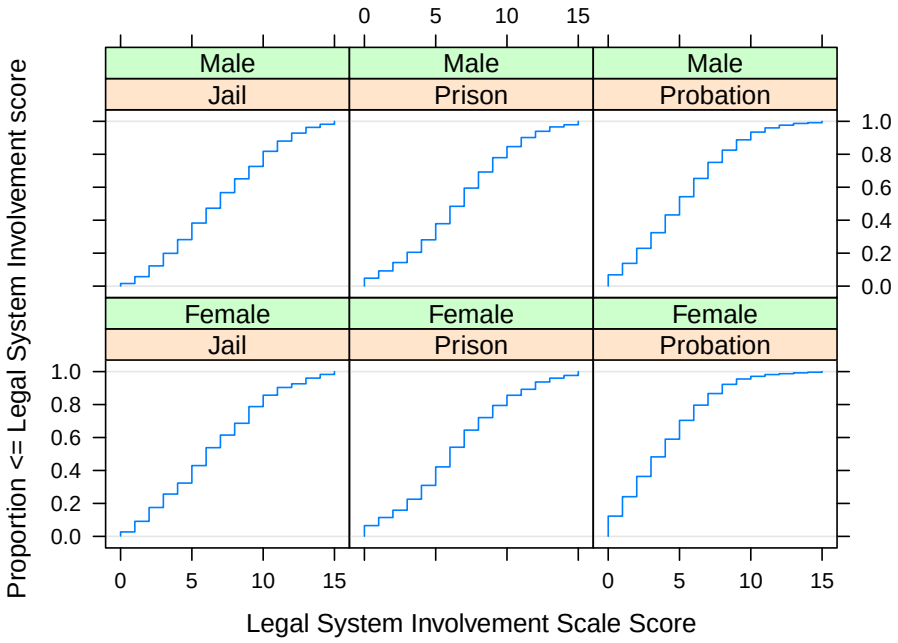


Figure 1: Cumulative distribution plots comparing subgroups in the norm data on the Legal System Involvement scale

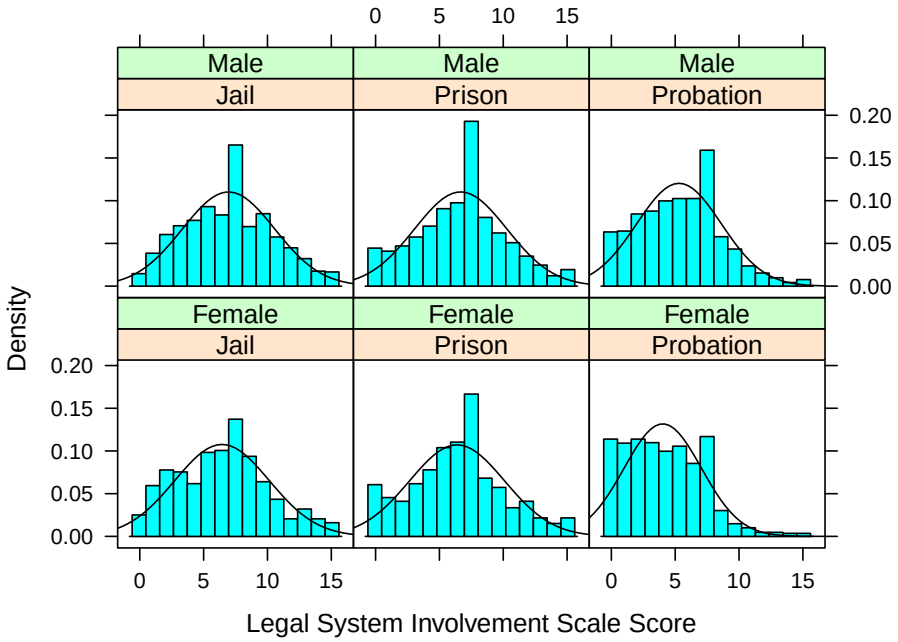


Figure 2: Histograms comparing subgroups in the norm data on the Legal System Involvement scale

The cumulative distributions for the subgroups are shown in Figure 1. These plots enhance the information in Table 8. For example, the plot in the lower right hand corner shows that female

probationers rarely scored above a 10 on the Legal System Involvement scale, compared to about 15% of female jail detainees scoring above 10 on this scale.

The histograms for the subgroups are shown in Figure 2. These plots demonstrate the tendency of lower scores among probationers, and somewhat lower scores among Women.

Score ranges

Tables 9 and 10 show the score ranges for each of the subgroups in the norm data. GN is a subgroup that includes all the Men and Women. Since there are more Men in the data, the GN ranges more strongly reflect the ranges for Men than for Women. Most agencies choose to use ranges based on gender-specific norms; GN is not available in the software at this time.

The ranges are determined by splitting the scores of each subgroup into the lowest 40% of scores for Low, the highest 30% of scores for High, and the remaining 30% in between for Medium. This table of ranges reflects what was seen earlier in the score distributions by subgroup. The lowest 40% of Women score have scores that are lower than the lowest 40% of Men. Similar differences between Men and Women are seen at the High level as well.

Table 9: Score ranges for the Legal System Involvement scale (standard version).

	Level		
	Low	Medium	High
GN Composite	0 – 5	6 – 8	9 – 15
GN Prison	0 – 6	7 – 9	10 – 15
GN Probation	0 – 4	5 – 7	8 – 15
GN Jail (Client-specific)	0 – 6	7 – 9	10 – 15
Male Composite	0 – 5	6 – 8	9 – 15
Male Prison	0 – 6	7 – 9	10 – 15
Male Probation	0 – 4	5 – 7	8 – 15
Male Jail (Client-specific)	0 – 6	7 – 9	10 – 15
Female Composite	0 – 4	5 – 7	8 – 15
Female Prison	0 – 5	6 – 8	9 – 15
Female Probation	0 – 3	4 – 5	6 – 15
Female Jail (Client-specific)	0 – 5	6 – 9	10 – 15

Scores from the Legal System Involvement scale (standard version) are used as inputs to the Summative GRRS, just as they are in the standard COMPAS Core GRRS. If a conviction version of the summative GRRS were to be developed at a later time, it is expected that scores from the conviction version of the Legal System Involvement scale would be used as inputs.

Table 10: Score ranges for the Legal System Involvement scale (conviction version).

	Level		
	Low	Medium	High
GN Composite	0 – 3	4 – 6	7 – 15
GN Prison	0 – 4	5 – 7	8 – 15
GN Probation	0 – 2	3 – 5	6 – 15
Male Composite	0 – 3	4 – 7	8 – 15
Male Prison	0 – 4	5 – 7	8 – 15
Male Probation	0 – 2	3 – 5	6 – 15
Female Composite	0 – 1	2 – 4	5 – 15
Female Prison	0 – 2	3 – 5	6 – 15
Female Probation	0 – 1	2 – 4	5 – 15

Mokken Scale Analysis (MSA): Checking the observable measurement properties of the Monotone Homogeneity Model

Item-pair scalability coefficient The final Legal System Involvement scale variables are listed in Table 3. Table 11 displays the item-pair scalability coefficients for all Legal System Involvement scale items. All values are positive. The smallest scalability coefficient (0.539) is between *n_prison and n_probations*. The largest scalability coefficient is 0.677 between *t_prev_arrests and n_probations*.

Table 11: Item-pair scalability coefficients (H_{ij}). Higher H_{ij} result in stronger scales; $-\infty < H_{ij} \leq 1$. The monotone homogeneity (IRT) model assumption requires that $0 \leq H_{ij} \leq 1$.

	t_prev_arrests	se	jail30	se	n_prison	se	n_probations	se
t_prev_arrests			0.573	(0.008)	0.604	(0.010)	0.677	(0.007)
jail30	0.573	(0.008)			0.635	(0.012)	0.604	(0.010)
n_prison	0.604	(0.010)	0.635	(0.012)			0.539	(0.008)
n_probations	0.677	(0.007)	0.604	(0.010)	0.539	(0.008)		

Item-scalability coefficient Table 12 displays the item scalability coefficient for all items in the Legal System Involvement scale. As desired, all coefficients are above the lower bound of $c = 0.30$. The scalability coefficient for the item *t_prev_arrests* actually decreased slightly from 0.670 prior to *jail30* being dichotomous. However, the remaining item scalability coefficients slightly increased after the transformation.

Table 12: Mokken Analysis: Item H values should be above 0.3

	Item H	se
t_prev_arrests	0.634	(0.006)
jail30	0.599	(0.007)
n_prison	0.570	(0.007)
n_probations	0.602	(0.005)

Test Scalability Coefficient The test scalability coefficient (H) for the Legal System Involvement scale in its final form is 0.601 which is within the limits for the monotone homogeneity model. A scalability coefficient of $H = 0.601$ implies that the scale is strong. The standard error for the statistic is (0.005). The original data had a test scalability coefficient of 0.591 with a standard error of (0.005).

Differential Item Functioning (DIF) Differential Item Functioning (DIF) was tested on a subset of the data (n = 1,000) after variable transformations. Table 13 denotes the ethnicity proportions from the random subset of n = 1,000 from the transformed data. No items were flagged in this subset; in the original data, where jail30 was not a dichotomous variable, a random subset of n = 1,000 was also used to test for DIF. The variables *n_prison and n_probations* were flagged. This does not imply that the transformation prevented those two items from being identified. Further analysis should be considered in the future; depending on the sample (n = 1,000), the items *n_prison* and *n_probations*, however, may indicate DIF. The subset of n = 1,000 were used to avoid small differences from being significant due to a large sample.

Table 13: Percent of Black, Hispanic, and White persons in a random sample of 1,000 cases drawn from the composite norm data; the data includes the variable n_jail after it being transformed (made dichotomous).

Ethnicity	Percent
Black	40
Hispanic	21
White	38

Non-Parametric Monotone Homogeneity (IRT) Model

The three assumptions for the monotone homogeneity model are the *Unidimensionality*, *Local Independence*, and *Latent Monotonicity*.

Unidimensionality The Unidimensionality assumption is tested using Principal Components Analysis (PCA). This was done in section 7.1. We see in Table 6 that the first eigenvalue (3.788) is much larger than the remaining eigenvalues and over sixty percent (63.4%) of the variance is explained. This is evidence that the *Legal System Problems* scale is unidimensional.

Local Independence The assumption of local independence was tested on the DIF sample of $n = 1{,}000$ using the Chi-Squared Test of Independence in R with the parameter `simulate.p.value` set to `TRUE`. All p-values were found to be less than 0.001, indicating evidence of local dependence.

Latent Monotonicity The latent monotonicity assumption is investigated by examining the observable property of the test data known as manifest monotonicity. We see in Table 14 that there are no violations (#vi).

Table 14: Manifest monotonicity results: used to assess latent monotonicity. The function `check.mononotinicity` was used from the R package *mokken*.

	ItemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
t_prev_arrests	0.63	84	0	0	0	0	0	0	0	0
jail30	0.60	21	0	0	0	0	0	0	0	0
n_prison	0.57	105	0	0	0	0	0	0	0	0
n_probations	0.60	99	0	0	0	0	0	0	0	0

7.2 History of Non-compliance Scale*

In the development sample, there were 12,425 individuals with completed History of Non-compliance scales whose responses were used to derive the standard version of the History of Non-compliance scale. 5,299 individuals had completed a conviction version History of Non-compliance scale.[2]

Review of Theoretical and Substantive Validity

This scale is an outcome variable and, as such, it is not expected to have complex substantive validity issues.

Scale Items

The standard and conviction versions of the History of Non-compliance scale are constructed using five items from the Criminal Involvement domain of the COMPAS-R Core. These items and their short descriptions are shown in Tables 15 and 16.

This scale measures on the degree to which the person has failed in the past when he or she has been placed in a community setting. Events considered include the number of times probation or parole has been suspended or revoked and the number of times a new charge or arrest has occurred while the person was on probation or parole.

The analyses that follow apply only to the standard version of the History of Non-compliance scale.

Table 15: COMPAS-R Core History of Non-compliance Scale Items (Standard Version). Practitioners are asked to verify all responses with the individual's criminal record. Any additional information that is shared by the individual that cannot be verified should be noted and scored. Include the current case for all questions.

Item	Short Description (Response Categories)
n_parole_return_R	How many times has this person been returned to custody for a parole violation? *(None=0; 1 time=1; 2 times=2; 3 times=3; 4 times=4; 5 or more times=5)*
n_parole_revo_R	How many times has this person's parole from prison been violated? *(None=0; 1 time=1; 2 times=2; 3 times=3; 4 times=4; 5 or more times=5)*
n_prob_arrest_R	How many times has this person had a new arrest or charge while on probation? *(None=0; 1 time=1; 2 or more times=2)*
n_prob_revo_R	How many times has this person's probation been violated or revoked? *(None=0; 1 time=1; 2 or more times=2)*
probpar_R	Was this person on probation or parole at the time of the current offense? *(Probation=1; Parole=1; Both=1; Neither=0)*

[2]Scales denoted with an asterisk (*) have both standard and conviction versions.

Table 16: COMPAS-R Core History of Non-compliance Scale Items (Conviction Version). Practitioners are asked to verify all responses with the individual's criminal record. Any additional information that is shared by the individual that cannot be verified should be noted and scored. Include the current case for all questions.

Item	Short Description (Response Categories)
n_parole_return_R	How many times has this person been returned to custody for a parole violation? *(None=0; 1 time=1; 2 times=2; 3 times=3; 4 times=4; 5 or more times=5)*
n_parole_revo_R	How many times has this person's parole from prison been violated? *(None=0; 1 time=1; 2 times=2; 3 times=3; 4 times=4; 5 or more times=5)*
n_prob_conv_R	How many times has this person had a new charge/arrest resulting in a conviction while on probation? *(None=0; 1 time=1; 2 or more times=2)*
n_prob_revo_R	How many times has this person's probation been violated or revoked? *(None=0; 1 time=1; 2 or more times=2)*
probpar_R	Was this person on probation or parole at the time of the current offense? *(Probation=1; Parole=1; Both=1; Neither=0)*

Item changes from the History of Non-compliance scale in the standard COMPAS Core

This scale contains the same five items as the corresponding scale in the standard COMPAS Core. However, various response configurations were tested to find a combination that simplified responses and was not detrimental to reliability. The changes made increased scale reliability from 0.687 to 0.731, putting it well into the acceptable range for scales within the social sciences (Cortina 1993).

Responses were changed on items pertaining to probation only:

- choices for the number of new arrests or charges while on probation were reduced to none, once, or more than once. The choices on the standard COMPAS Core are None; 1 time; 2 times; 3 times; 4 times; and 5 or more times.
- choices for the number of times the person's probation had been violated or revoked were reduced to none, once, or more than once. The choices on the standard COMPAS Core are None; 1 time; 2 times; 3 times; 4 times; and 5 or more times.

Additional statistics for the scale are included in this section.

Factor structure in the norm data

Table 17: Principal component loadings.

	PC1	PC2	PC3	PC4	PC5
n_parole_return	0.285	-0.548	0.033	-0.016	-0.785
n_parole_revo	0.376	-0.688	0.043	0.004	0.619
n_prob_arrest	0.573	0.347	0.684	0.288	-0.012
n_prob_revo	0.636	0.297	-0.709	0.060	-0.007
probpar	0.209	0.130	0.161	-0.956	0.012

Table 18: Eigenvalues and the proportion of the variance explained by each principal component.

	PC1	PC2	PC3	PC4	PC5
Eigenvalue	1.251	0.662	0.208	0.142	0.069
Standard deviation	1.118	0.814	0.456	0.377	0.263
Proportion of variance	0.536	0.284	0.089	0.061	0.030
Cumulative proportion	0.536	0.820	0.909	0.970	1.000

Principal components analysis of the History of Non-compliance items in the COMPAS-R composite norm group is presented here. The loadings of the principal components are shown in Table 17. The most important loadings on PC1 are the probation items, while the most important loadings on PC2 are the parole items.

Table 18 shows the eigenvalues of the principal components and the proportion of the variance explained by each. As shown in the table, the first principal component has an eigenvalue of 1.251 and explains 53.6% of the variance in the data. Over 90% of the variance is explained by the first three components.

Scale consistency and reliability in the norm data

Table 19: Correlation of items with scale score, and alpha reliability of the scale when each item is removed. Compare with alpha of 0.731 for all five items.

	Correlation with score	Alpha when item removed
n_parole_return	0.663	0.690
n_parole_revo	0.694	0.691
probpar	0.561	0.719
n_prob_arrest	0.757	0.662
n_prob_revo	0.788	0.648

Cronbach's alpha measures a scale's internal consistency and reliability. Alpha for the History of Non-compliance scale in the composite norm set is 0.731. Table 19 shows the correlation of each item with the overall scale score, and the value of alpha when the item is removed from the scale. Removing any of the items would substantially lower alpha.

Scale distributions in the norm groups

Table 20: Tabulation by population and gender with mean scores. Note: this table does not present the scale score from the single transgender person in the data.

Group	N	Mean
Jail Female	405	1.8
Prison Female	856	2.2
Probation Female	1561	1.0
Jail Male	1900	2.1
Prison Male	3350	2.5
Probation Male	4353	1.5
ALL	12425	1.9

Table 20 shows the mean History of Non-compliance scale score for each population subgroup in the normative data. The relationships among these means are as expected, with Women generally lower when comparing scores by gender, and those on probation generally lower when comparing correction types.

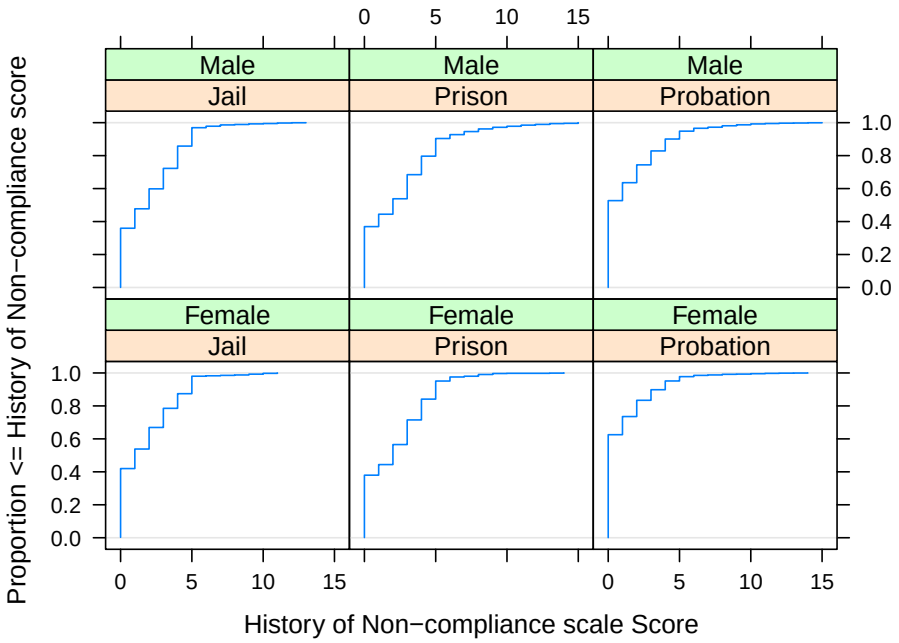


Figure 3: Cumulative distribution plots comparing subgroups in the norm data on the History of Non-compliance scale

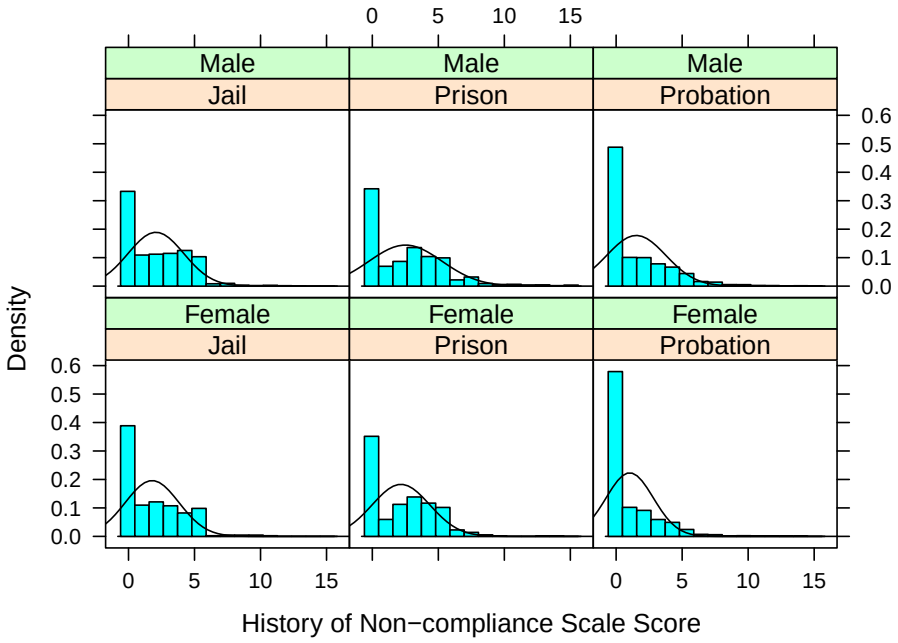


Figure 4: Histograms comparing subgroups in the norm data on the History of Non-compliance scale

The cumulative distributions for the subgroups are shown in Figure 3. These plots enhance the information in Table 20. For example, the plot in the lower right hand corner shows that about 60% of female probationers scored 0 on the History of Non-compliance scale, compared to only about 40% of female jail or prison detainees.

The histograms for the subgroups are shown in Figure 4. These plots demonstrate the tendency of lower scores among probationers, and somewhat lower scores among Women.

Score ranges

Tables 21 and 22 show the score ranges for each of the subgroups in the norm data, for the standard and conviction versions of the History of Non-compliance scale, respectively. GN is a subgroup that includes all the Men and Women. Since there are more Men in the data, the GN ranges more strongly reflect the ranges for Men than for Women. Most agencies choose to use ranges based on gender-specific norms; GN is not available in the software at this time.

The ranges are determined by splitting the scores of each subgroup into the lowest 40% of scores for Low, the highest 30% of scores for High, and the remaining 30% in between for Medium.

Table 21: Score ranges for the History of Non-compliance scale (standard version).

	Level		
	Low	Medium	High
GN Composite	0	1 – 3	4 – 15
GN Prison	0 – 1	2 – 4	5 – 15
GN Probation	0	1 – 2	3 – 15
GN Jail (Client-specific)	0 – 1	2 – 3	4 – 15
Male Composite	0	1 – 3	4 – 15
Male Prison	0 – 1	2 – 4	5 – 15
Male Probation	0	1 – 2	3 – 15
Male Jail (Client-specific)	0 – 1	2 – 3	4 – 15
Female Composite	0	1 – 2	3 – 15
Female Prison	0 – 1	2 – 3	4 – 15
Female Probation	0	1	2 – 15
Female Jail (Client-specific)	0	1 – 3	4 – 15

Table 22: Score ranges for the History of Non-compliance scale (conviction version).

	Level		
	Low	Medium	High
GN Composite	0	1 – 2	3 – 15
GN Prison	0	1 – 3	4 – 15
GN Probation	0	1	2 – 15
Male Composite	0	1 – 2	3 – 15
Male Prison	0 – 1	2 – 4	5 – 15
Male Probation	0	1	2 – 15
Female Composite	0		1 – 15
Female Prison	0	1 – 2	3 – 15
Female Probation	0		1 – 15

Mokken Scale Analysis (MSA): Checking the observable measurement properties of the Monotone Homogeneity Model

Item-pair scalability coefficient The final History of Non-Compliance scale variables are listed in Table 15. Table 23 displays the item-pair scalability coefficients for all History of Non-Compliance scale items. All values are positive. The smallest scalability coefficient (0.267) is between *probpar and n_parole_revo*. The largest scalability coefficient is 0.874 between *n_parole_revo and n_parole_return*.

Table 23: Item-pair scalability coefficients (H_{ij}). Higher H_{ij} result in stronger scales; $-\infty < H_{ij} \leq 1$. The monotone homogeneity (IRT) model assumption requires that $0 \leq H_{ij} \leq 1$.

	probpar	n_parole_revo	n_parole_return	n_prob_arrest	n_prob_revo
probpar		0.267	0.274	0.540	0.508
n_parole_revo	0.267		0.874	0.380	0.563
n_parole_return	0.274	0.874		0.374	0.570
n_prob_arrest	0.540	0.380	0.374		0.740
n_prob_revo	0.508	0.563	0.570	0.740	

* The standard errors (se) have been removed to allow all item names to fit in the table. The standard errors range from 0.007 to 0.024. The smallest item-pair scalability coefficient is 0.267 and the largest is 0.874.

Item-scalability coefficient The item scalability coefficients are in Table 24; they are all above the lower bound of 0.3 as desired.

Table 24: Mokken Analysis: Item H values should be above 0.3

	Item H	se
probpar	0.448	(0.010)
n_parole_revo	0.583	(0.011)
n_parole_return	0.619	(0.011)
n_prob_arrest	0.561	(0.009)
n_prob_revo	0.625	(0.007)

Differential Item Functioning (DIF) Differential Item Functioning (DIF) was tested on a subset of the data ($n = 1{,}000$) for race, after variable transformations. Table 25 denotes the ethnicity proportions from the random subset of $n = 1{,}000$ from the transformed data. All items were flagged. In the original data, prior to modifying the items: *n_prob_arrest* and *n_prob_revo*, a random subset of $n = 1{,}000$ was also used to test for DIF. All items were flagged.

Non-Parametric Monotone Homogeneity (IRT) Model

The three assumptions for the monotone homogeneity model are the *Unidimensionality*, *Local Independence*, and *Latent Monotonicity*.

Unidimensionality The Unidimensionality assumption is tested using Principal Components Analysis (PCA). This was done in section 7.2. We see in Table 18 that the first eigenvalue (1.251)

Table 25: Percent of Black, Hispanic, and White persons in a random sample of 1,000 cases drawn from the composite norm data; All variable transformations have been incorporated.

Ethnicity	Percent
Black	43
Hispanic	23
White	34

is almost twice as large as the second eigenvalue. Furthermore, 53.6% of the variance is explained in the first principal component. This is evidence that the *History of Non-Compliance* scale is unidimensional.

Local Independence The assumption of local independence was tested on the DIF sample of $n = 1{,}000$ using the Chi-Squared Test of Independence in R with the parameter `simulate.p.value` set to `TRUE`. All p-values were found to be less than 0.001, indicating evidence of local dependence.

Latent Monotonicity The latent monotonicity assumption is investigated by examining the observable property of the test data known as manifest monotonicity. We see in Table 26 that there are no violations (#vi).

Table 26: Manifest monotonicity results: used to assess latent monotonicity (3rd assumtion for monotone homogeneity (IRT) model. The function `check.mononotinicity` was used from the R package *mokken*.

	ItemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
probpar	0.45	3	0	0	0	0	0	0	0	0
n_parole_revo	0.58	50	0	0	0	0	0	0	0	0
n_parole_return	0.62	24	0	0	0	0	0	0	0	0
n_prob_arrest	0.56	12	0	0	0	0	0	0	0	0
n_prob_revo	0.63	12	0	0	0	0	0	0	0	0

7.3 History of Violence Scale*

Both the standard and conviction versions of the History of Violence scale are derived from the respective standard and conviction versions of the scale in the standard COMPAS Core.[3] In the development sample, there were 15,731 individuals with completed History of Violence scales that were used to derive the standard version of the History of Violence scale. 5,299 individuals had completed a conviction version of the History of Violence scale.

Review of Theoretical and Substantive Validity

A history of violent behavior, either in adolescence or adulthood, is one of the best predictors of future violence (Farrington 1991; Parker and Asher 1987).

Scale Items

The standard and conviction versions of the History of Violence scale are constructed using six items from the Criminal Involvement domain of the COMPAS-R Core. These items and their short descriptions are shown in Tables 27 and 28.

As in the standard COMPAS Core, the purpose of this scale is to measure the seriousness and extent of violence in an person's criminal history. Central to this is the frequency with which a person has committed violent felony offenses. The use of weapons in the commission of violent crimes and the seriousness of the charges are also considered. The frequency of violent offenses is broken down by categories such as domestic violence and violent misdemeanor or felony offenses.

The analyses that follow apply only to the standard version of the History of Violence scale.

[3] Scales denoted with an asterisk (*) have both standard and conviction versions.

Table 27: COMPAS-R Core History of Violence Scale Items (Standard Version). Practitioners are asked to verify all responses with the individual's criminal record. Any additional information that is shared by the individual that cannot be verified should be noted and scored. Exclude the current case for all questions.

Item	Short Description (Response Categories)
domviol_R	How many prior family violence offense arrests as an adult? *(0=0; 1=1; 2 or more=2)*
fights_inmate_R	Has this person, while incarcerated in jail or prison, ever received serious or administrative disciplinary infractions for fighting/threatening other inmates or staff? *(No=0; Yes=1)*
n_fel_assault_R	How many prior felony assault offense arrests, including those involving sex offense (with force) and homicide/voluntary manslaughter, as an adult? Exclude domestic violence. *(0=0; 1=1; 2=2; 3=3; 4 or more=4)*
n_misd_assault_R	How many prior misdemeanor assault offense arrests, including those involving sex offense (with force) and manslaughter, as an adult? Exclude domestic violence. *(0=0; 1=1; 2 or more=2)*
n_violent_property_R	How many prior arrests for a felony property offense that included an element of violence as an adult? *(0=0; 1=1; 2 or more=2)*
weapons_offense_R	How many prior weapons offense arrests as an adult? *(0=0; 1=1; 2 or more=2)*

Table 28: COMPAS-R Core History of Violence Scale Items (Conviction Version). Practitioners are asked to verify all responses with the individual's criminal record. Any additional information that is shared by the individual that cannot be verified should be noted and scored. Exclude the current case for all questions.

Item	Short Description (Response Categories)
domviolconv_R	How many prior family violence offense convictions as an adult? *(0=0; 1=1; 2 or more=2)*
fights_inmate_R	Has this person, while incarcerated in jail or prison, ever received serious or administrative disciplinary infractions for fighting/threatening other inmates or staff? *(No=0; Yes=1)*
n_fel_assaultconv_R	How many prior felony assault offense convictions, including those involving sex offense (with force) and homicide/voluntary manslaughter, as an adult? Exclude domestic violence. *(0=0; 1=1; 2=2; 3=3; 4 or more=4)*
n_misd_assaultconv_R	How many prior misdemeanor assault offense convictions, including those involving sex offense (with force) and manslaughter, as an adult? Exclude domestic violence *(0=0; 1=1; 2 or more=2)*
n_violent_propconv_R	How many prior convictions for a felony property offense that included an element of violence as an adult? *(0=0; 1=1; 2 or more=2)*
weapons_offenseconv_R	How many prior weapons offense convictions as an adult? *(0=0; 1=1; 2 or more=2)*

Item changes from the History of Violence scale in the standard COMPAS Core

Items that have fewer response choices in COMPAS-R Core than standard COMPAS Core are

- number of felony property offenses that included an element of violence,
- number of misdemeanor assaults,
- number of weapons offenses, and
- number of domestic violence offenses.

The item asking for number of felony assault offenses has an *increased* number of responses since it now includes homicide/manslaughter and sex offenses with force.

The item asking for number of juvenile violent felony offenses has been removed because it is inconsistently available. Its removal has little impact on the scale performance.

The item asking whether the person ever received infractions for fighting or threatening inmates or staff during a prior incarceration is unchanged.

These changes collectively increased the reliability of the scale modestly from 0.617 to 0.626. Additional statistics for the scale are included in this section.

Factor structure in the norm data

Table 29: Principal component loadings.

	PC1	PC2	PC3	PC4	PC5	PC6
domviol	-0.409	0.587	-0.159	0.679	-0.046	-0.008
fights_inmate	-0.081	-0.059	0.025	-0.003	0.009	-0.995
n_fel_assault	-0.614	-0.526	-0.553	-0.056	-0.185	0.065
n_misd_assault	-0.515	0.483	0.179	-0.684	0.016	0.020
n_violent_property	-0.317	-0.299	0.762	0.219	-0.421	0.058
weapons_offense	-0.290	-0.230	0.235	0.139	0.886	0.050

Table 30: Eigenvalues and the proportion of the variance explained by each principal component.

	PC1	PC2	PC3	PC4	PC5	PC6
Eigenvalue	0.962	0.430	0.328	0.308	0.244	0.089
Standard deviation	0.981	0.656	0.573	0.555	0.494	0.299
Proportion of variance	0.407	0.182	0.139	0.130	0.103	0.038
Cumulative proportion	0.407	0.590	0.729	0.859	0.962	1.000

Principal components analysis of the History of Violence items in the COMPAS-R composite norm group is presented here. The loadings of the principal components are shown in Table 29. Table 30 shows the eigenvalues of the principal components and the proportion of the variance explained by each. As shown in the table, the first principal component has an eigenvalue of 0.962 and explains 40.7% of the variance in the data. The first five principal components are needed to explain over 90% of the variance in the data.

Scale consistency and reliability in the norm data

Table 31: Correlation of items with scale score, and alpha reliability of the scale when each item is removed. Compare with alpha of 0.626 for all six items.

	Correlation with score	Alpha when item removed
domviol	0.610	0.591
fights_inmate	0.365	0.625
n_fel_assault	0.669	0.538
n_misd_assault	0.681	0.555
n_violent_property	0.588	0.589
weapons_offense	0.587	0.579

Cronbach's alpha measures a scale's internal consistency and reliability. Alpha for the History of Violence scale in the composite norm set is 0.626. Table 31 shows the correlation of each item with the overall scale score, and the value of alpha when the item is removed from the scale. Removing

the fighting while incarcerated item affects alpha minimally, but it is an important item to have information about, and it is well-correlated with the scale score. Removing any of the other items would lower alpha more substantially.

Scale distributions in the norm groups

Table 32: Tabulation by population and gender with mean scores. Note: this table does not present the scale score from the single transgender person in the data.

Group	N	Mean
Jail Female	405	1.9
Prison Female	1015	1.3
Probation Female	2091	0.8
Jail Male	1900	3.1
Prison Male	4043	2.2
Probation Male	6276	1.6
ALL	15730	1.8

Table 32 shows the mean History of Violence scale score for each population subgroup in the normative data. The relationships among these means are as expected, with Women substantially lower than Men when comparing scores by gender, and those on probation lower when comparing correction settings.

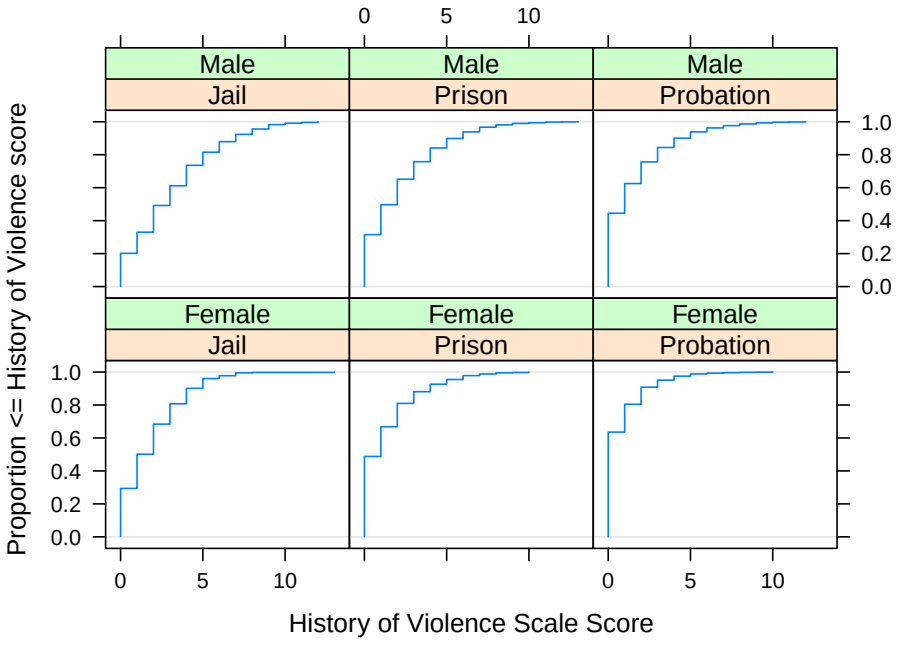


Figure 5: Cumulative distribution plots comparing subgroups in the norm data on the History of Violence scale

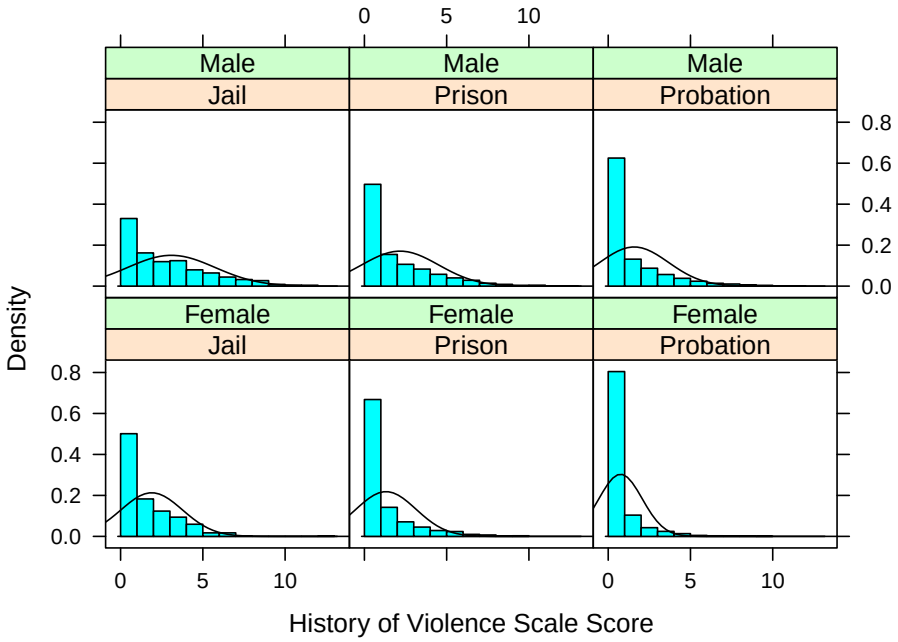


Figure 6: Histograms comparing subgroups in the norm data on the History of Violence scale

The cumulative distributions for the subgroups are shown in Figure 5. These plots enhance the information in Table 32. For example, the plot in the lower right hand corner shows that the majority of female probationers scored a 0 on the History of Violence scale, compared to only about 30% of female jail detainees.

The histograms for the subgroups are shown in Figure 6. These plots demonstrate the tendency of lower scores among probationers, and somewhat lower scores among Women.

Score ranges

Tables 33 and 34 show the score ranges for each of the subgroups in the norm data. GN is a subgroup that includes all the Men and Women. Since there are more Men in the data, the GN ranges more strongly reflect the ranges for Men than for Women. Most agencies choose to use ranges based on gender-specific norms; GN is not available in the software at this time.

The ranges for both the standard and conviction versions of the scales are determined by splitting the scores of each subgroup into the lowest 40% of scores for Low, the highest 30% of scores for High, and the remaining 30% in between for Medium. This table of ranges reflects what was seen earlier in the score distributions by subgroup.

Table 33: Score ranges for the History of Violence scale (standard version).

	Level		
	Low	Medium	High
GN Composite	0	1 – 2	3 – 13
GN Prison	0 – 1	2 – 3	4 – 13
GN Probation	0	1 – 2	3 – 13
GN Jail (Client-specific)	0 – 2	3 – 4	5 – 13
Male Composite	0 – 1	2 – 3	4 – 13
Male Prison	0 – 1	2 – 3	4 – 13
Male Probation	0	1 – 2	3 – 13
Male Jail (Client-specific)	0 – 2	3 – 4	5 – 13
Female Composite	0	1	2 – 13
Female Prison	0	1 – 2	3 – 13
Female Probation	0	1	2 – 13
Female Jail (Client-specific)	0 – 1	2 – 3	4 – 13

Table 34: Score ranges for the History of Violence scale (conviction version).

	Level		
	Low	Medium	High
GN Composite	0	1	2 – 13
GN Prison	0	1 – 2	3 – 13
GN Probation	0		1 – 13
Male Composite	0	1	2 – 13
Male Prison	0 – 1	2	3 – 13
Male Probation	0	1	2 – 13
Female Composite	0		1 – 13
Female Prison	0		1 – 13
Female Probation	0		1 – 13

Mokken Scale Analysis (MSA): Checking the observable measurement properties of the Monotone Homogeneity Model

Item-pair scalability coefficient The final History of Violence scale variables are listed in Table 27. Table 35 displays the item-pair scalability coefficients for all History of Violence scale items. All values are positive with the smallest scalability coefficient (0.088) between *domviol and fights_inmate*. The largest scalability coefficient is 0.376 between *n_misd_assault and domviol*.

Table 35: Item-pair scalability coefficients (H_{ij}). Higher H_{ij} result in stronger scales; $-\infty < H_{ij} \leq 1$. The monotone homogeneity (IRT) model assumption requires that $0 \leq H_{ij} \leq 1$.

	n_violent_property	n_misd_assault	n_fel_assault	domviol	weapons_offense	fights_inmate
n_violent_property		0.237	0.296	0.154	0.283	0.193
n_misd_assault	0.237		0.372	0.376	0.242	0.165
n_fel_assault	0.296	0.372		0.273	0.342	0.251
domviol	0.154	0.376	0.273		0.184	0.088
weapons_offense	0.283	0.242	0.342	0.184		0.191
fights_inmate	0.193	0.165	0.251	0.088	0.191	

* The standard errors (se) have been removed to allow all item names to fit in the table. The standard errors range from 0.01 to 0.013. The smallest item-pair scalability coefficient is 0.088 and the largest is 0.376.

Item-scalability coefficient The item scalability coefficients are in Table 36; not all items are above the desired lower bound of 0.3.

Table 36: Mokken Analysis: Item H values should be above 0.3

	Item H	se
n_violent_property	0.238	(0.006)
n_misd_assault	0.300	(0.007)
n_fel_assault	0.313	(0.006)
domviol	0.239	(0.007)
weapons_offense	0.259	(0.007)
fights_inmate	0.179	(0.007)

Test Scalability Coefficient The test scalability coefficient (H) for the History of Violence scale in its final form is 0.263 which is within the limits for the monotone homogeneity model. A scalability coefficient of $H = 0.263$ implies that the scale is very weak. The standard error for the statistic is (0.005). The original data had a test scalability coefficient of 0.222 with a standard error of (0.004).

Differential Item Functioning (DIF) Differential Item Functioning (DIF) was tested on a subset of the data ($n = 1{,}000$) for race, after variable transformations. The *Other* category for race was not included since it only made up 7.8% of the final sample. Table 37 denotes the ethnicity proportions from the random subset of $n = 1{,}000$ from the transformed data.

The following items were found in the random subsample of $n = 1{,}000$ to have differential item functioning (DIF): *n_misd_assault, domviol, weapons_offense and fights_inmate*. In the original data, a random subsample of $n = 1{,}000$ was also selected; the following variables were flagged with DIF: *homicide, n_misd_assault, n_fel_assault, juv_vfelony, domviol and fights_inmate*.

Table 37: Percent of Black, Hispanic, and White persons in a random sample of 1,000 cases drawn from the composite norm data; the data includes all variable transformations.

Ethnicity	Percent
Black	40
Hispanic	23
White	38

Non-Parametric Monotone Homogeneity (IRT) Model

The three assumptions for the monotone homogeneity model are the *Unidimensionality*, *Local Independence*, and *Latent Monotonicity*.

Unidimensionality The Unidimensionality assumption is tested using Principal Components Analysis (PCA). This was done in section 7.3. We see in Table 30 that the first eigenvalue (0.962) is larger than the remaining eigenvalues, but only a little over 40 percent (40.7%) of the variance is explained. This is moderate evidence that the History of Violence scale is unidimensional.

Local Independence The assumption of local independence was tested using the Chi-Squared Test of Independence in R with the parameter `simulate.p.value` set to `TRUE`. The $n = 1{,}000$ subset used for testing DIF from the transformed data (items changed for the COMPAS-R Core scales) was used. There was 1 pair of items with p-values greater than or equal to 0.05. These item pairs are listed in Table 38.

Table 38: Item pair and its corresponding p-value for items that suggest independence from a subset of n = 1,000 from the final COMPAS-R CORE scale.

item1	item2	p_value
domviol	fights_inmate	0.698

Latent Monotonicity The latent monotonicity assumption is investigated by examining the observable property of the test data known as manifest monotonicity. We see in Table 39 that there are no violations (#vi).

Table 39: Manifest monotonicity results: used to assess latent monotonicity. The function `check.mononotinicity` was used from the R package *mokken*.

	ItemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
n_violent_property	0.24	12	0	0	0	0	0	0	0	0
n_misd_assault	0.30	12	0	0	0	0	0	0	0	0
n_fel_assault	0.31	24	0	0	0	0	0	0	0	0
domviol	0.24	12	0	0	0	0	0	0	0	0
weapons_offense	0.26	12	0	0	0	0	0	0	0	0
fights_inmate	0.18	10	0	0	0	0	0	0	0	0

7.4 Current Violence Scale

The Current Violence scale of the COMPAS-R Core is derived from the corresponding Current Violence scale in the standard COMPAS Core. In the development sample, there were 16,011 individuals with completed Current Violence scales.

Review of Theoretical and Substantive Validity

This scale is part of a person's criminal history. Its theoretical and substantive validity are of a piece with comments on criminal violence made elsewhere in this document and other COMPAS documents describing the psychometric properties of the COMPAS scales (e.g., Brennan, Dieterich, and Oliver (2007)).

Scale Items

The Current Violence scale is constructed using seven items from the Criminal Involvement domain of the COMPAS-R Core. These items and their short descriptions are shown in Table 40.

This scale measures the degree of violence in the present offense or charge. The scale is, in part, an inventory of the violent offenses the person is charged with as part of the current arrest. Additional information includes whether the present offense is an assaultive felony and whether any of the current charges involve domestic violence.

Table 40: COMPAS-R Core Current Violence Scale Items. Practitioners are instructed to verify all responses with the individual's criminal record. Any additional information that is shared by the individual that cannot be verified should be noted and scored.

Item	Short Description (Response Categories)
crime_cat_R	Which category represents the most serious current charge? *(Misdemeanor=0; Non-violent felony=0; Violent felony=1)*
curr_assault_R	Assault *(Not checked in Current Charges table=0; Checked in Current Charges table=1)*
curr_domviol_R	Do any current charges involve family violence? *(No=0; Yes=1)*
curr_homicide_R	Homicide *(Not checked in Current Charges table=0; Checked in Current Charges table=1)*
curr_robbery_R	Robbery *(Not checked in Current Charges table=0; Checked in Current Charges table=1)*
curr_sexwithforce_R	Sex Offense with Force *(Not checked in Current Charges table=0; Checked in Current Charges table=1)*
curr_weapons_R	Weapons *(Not checked in Current Charges table=0; Checked in Current Charges table=1)*

Item changes from the Current Violence scale in the standard COMPAS Core

No items were removed or changed in this scale for the COMPAS-R Core. The data with the complete set of items showed a weak alpha (0.52) that, nevertheless, did not improve by removing or altering items. Additional statistics for the scale are included in this section.

Factor structure in the norm data

Table 41: Principal component loadings.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
crime_cat	0.770	-0.455	0.277	-0.069	0.070	0.220	-0.256
curr_assault	0.514	0.552	-0.338	-0.474	0.124	-0.189	0.201
curr_domviol	0.317	0.505	0.092	0.740	-0.283	0.063	-0.068
curr_homicide	0.058	-0.092	0.031	0.070	0.015	0.547	0.826
curr_robbery	0.122	-0.292	-0.069	-0.093	-0.804	-0.415	0.259
curr_sexwithforce	0.099	-0.123	0.341	0.265	0.461	-0.660	0.373
curr_weapons	0.116	-0.353	-0.824	0.374	0.200	-0.060	-0.012

Table 42: Eigenvalues and the proportion of the variance explained by each principal component.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Eigenvalue	0.283	0.142	0.093	0.072	0.049	0.036	0.021
Standard deviation	0.532	0.377	0.304	0.268	0.221	0.189	0.143
Proportion of variance	0.407	0.205	0.133	0.104	0.070	0.051	0.030
Cumulative proportion	0.407	0.612	0.745	0.849	0.919	0.970	1.000

Principal components analysis of the Current Violence items in the COMPAS-R composite norm group is presented here. The loadings of the principal components are shown in Table 41. Table 42 shows the eigenvalues of the principal components and the proportion of the variance explained by each. As shown in the table, the first principal component has an eigenvalue of 0.283 and explains 40.7% of the variance in the data. Five principal components are required to explain at least 90% of the variation in the data.

Scale consistency and reliability in the norm data

Table 43: Correlation of items with scale score, and alpha reliability of the scale when each item is removed. Compare with alpha of 0.52 for all seven items.

	Correlation with score	Alpha when item removed
crime_cat	0.829	0.250
curr_assault	0.617	0.446
curr_domviol	0.540	0.470
curr_homicide	0.277	0.520
curr_robbery	0.375	0.510
curr_sexwithforce	0.311	0.525
curr_weapons	0.391	0.534

Cronbach's alpha measures a scale's internal consistency and reliability. Alpha for the Current Violence scale in the composite norm set is 0.52. Table 43 shows the correlation of each item with the overall scale score, and the value of alpha when the item is removed from the scale.

Removing the current weapons or sex with force offense item affects alpha minimally, but it was decided that the details of the violent offense were important to inventory using this scale. They are both well-correlated with the scale score. Removing any of the other items would simply lower alpha.

Scale distributions in the norm groups

Table 44: Tabulation by population and gender with mean scores. Note: this table does not present the scale score from the single transgender person in the data.

Group	N	Mean
Jail Female	404	0.5
Prison Female	956	0.7
Probation Female	1873	0.4
Jail Male	1897	0.9
Prison Male	4864	1.2
Probation Male	6016	0.7
ALL	16010	0.8

Table 44 shows the mean Current Violence scale score for each population subgroup in the normative data. The relationships among these means are as expected, with Women generally lower when comparing scores by gender, and those in prison generally higher when comparing correction types.

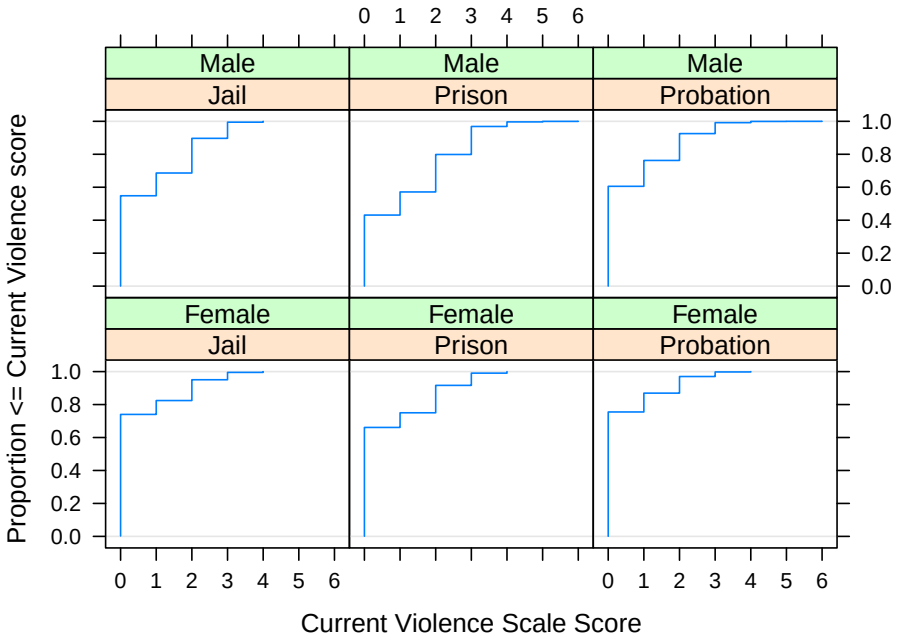


Figure 7: Cumulative distribution plots comparing subgroups in the norm data on the Current Violence scale

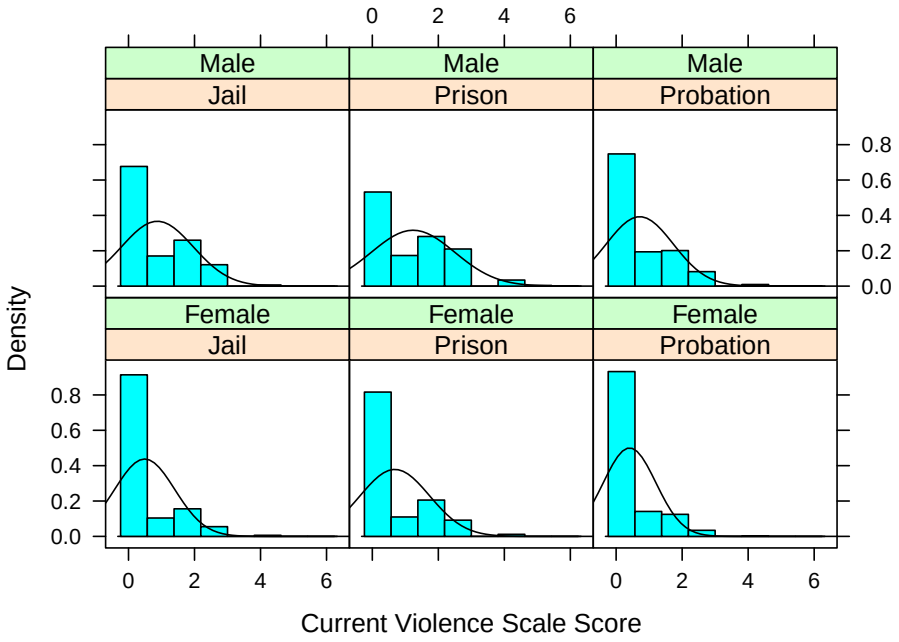


Figure 8: Histograms comparing subgroups in the norm data on the Current Violence scale

The cumulative distributions for the subgroups are shown in Figure 7 and histograms are shown in Figure 8. These plots enhance the information in Table 44. For example, most individuals in any

Table 45: Score ranges for the Current Violence scale

	Level		
	Low	Medium	High
GN Composite	0	1	2 – 7
GN Prison	0	1 – 2	3 – 7
GN Probation	0	1	2 – 7
GN Jail (Client-specific)	0	1	2 – 7
Male Composite	0	1 – 2	3 – 7
Male Prison	0	1 – 2	3 – 7
Male Probation	0	1	2 – 7
Male Jail (Client-specific)	0	1 – 2	3 – 7
Female Composite	0		1 – 7
Female Prison	0	1	2 – 7
Female Probation	0		1 – 7
Female Jail (Client-specific)	0		1 – 7

subgroup do not score above zero on this scale, with the exception of male prison inmates. They are the only group with a mean score greater than 1. The histograms make it clear that more Women score 0 on this assessment than Men, regardless of subgroup.

Score ranges

Table 45 shows the score ranges for each of the subgroups in the norm data. GN is a subgroup that includes all the Men and Women. Since there are more Men in the data, the GN ranges more strongly reflect the ranges for Men than for Women. Most agencies choose to use ranges based on gender-specific norms; GN is not available in the software at this time.

The ranges are determined by splitting the scores of each subgroup into the lowest 40% of scores for Low, the highest 30% of scores for High, and the remaining 30% in between for Medium. This table of ranges reflects what was seen earlier in the score distributions by subgroup. Since such large proportions of Women score 0, the middle 30% is also considered low, with the exception of Women prison inmates.

Mokken Scale Analysis (MSA): Checking the observable measurement properties of the Monotone Homogeneity Model

The Current Violence scale variables are listed in Table 40. The items are identical to the original *Current Violence* scale in COMPAS Core. The scoring was modified such that all scores start at zero, but no other item score modifications were done. MSA analysis was performed on the data and the results are provided below.

Item-pair scalability coefficient The final Current Violence scale variables are listed in Table 40. Table 46 displays the item-pair scalability coefficients for all History of Violence scale items. All values are not positive with the smallest scalability coefficient (-0.143) between *curr_assault and curr_sexwithforce*. The largest scalability coefficient is 0.959 between *crime_cat and curr_homicide*.

Table 46: Item-pair scalability coefficients (H_{ij}). Higher H_{ij} result in stronger scales; $-\infty < H_{ij} \leq 1$. The monotone homogeneity (IRT) model assumption requires that $0 \leq H_{ij} \leq 1$.

	crime_cat	curr_assault	curr_sexwithforce	curr_homicide	curr_robbery	curr_weapons	curr_domviol
crime_cat		0.533	0.853	0.959	0.866	0.333	0.397
curr_assault	0.533		-0.143	-0.103	-0.018	0.042	0.542
curr_sexwithforce	0.853	-0.143		-0.025	-0.037	-0.078	0.191
curr_homicide	0.959	-0.103	-0.025		0.033	0.193	0.097
curr_robbery	0.866	-0.018	-0.037	0.033		0.268	-0.109
curr_weapons	0.333	0.042	-0.078	0.193	0.268		-0.053
curr_domviol	0.397	0.542	0.191	0.097	-0.109	-0.053	

* The standard errors (se) have been removed to allow all item names to fit in the table. The standard errors range from 0.006 to 0.023. The smallest item-pair scalability coefficient is -0.143 and the largest is 0.959.

Item-scalability coefficient The item scalability coefficients are in Table 47; not all items are above the desired lower bound of 0.3.

Table 47: Mokken Analysis: Item H values should be above 0.3

	Item H	se
crime_cat	0.551	(0.008)
curr_assault	0.271	(0.007)
curr_sexwithforce	0.113	(0.005)
curr_homicide	0.165	(0.007)
curr_robbery	0.159	(0.006)
curr_weapons	0.103	(0.007)
curr_domviol	0.222	(0.006)

Test Scalability Coefficient The test scalability coefficient (H) for the Current Violence scale in its final form is 0.245 which is within the limits for the monotone homogeneity model. A scalablility coefficient of $H = 0.245$ implies that the scale is very weak. The standard error for the statistic is (0.004). The original data had a test scalability coefficient of 0.245 with a standard error of (0.004).

Table 48: Percent of Black, Hispanic, and White persons in a random sample of 1,000 cases drawn from the composite norm data; the data includes all variable transformations.

Ethnicity	Percent
Black	44
Hispanic	15
White	41

Differential Item Functioning (DIF) Differential Item Functioning (DIF) was tested on a subset of the data (n = 2,000)[4] for race, after variable transformations; the *Other* category for race was not included since it only made up 7.5% of the final sample. Table 48 denotes the ethnicity proportions from the random subset of n = 2,000 from the transformed data.

The following items were found in the random subsample of n = 2,000 to have differential item functioning (DIF): *curr_robbery, curr_weapons and curr_domviol*. In the original data, a random subsample of n = 2,000 was also selected; the following variables were flagged with DIF: *curr_sexwithforce, curr_robbery, curr_weapons and curr_domviol*.

Non-Parametric Monotone Homogeneity (IRT) Model

The three assumptions for the monotone homogeneity model are the *Unidimensionality*, *Local Independence*, and *Latent Monotonicity*.

Unidimensionality The Unidimensionality assumption is tested using Principal Components Analysis (PCA). This was done in section 7.4. We see in Table 42 that the first eigenvalue (0.283) is somewhat larger than the remaining eigenvalues, with only a little over 40 percent (40.7%) of the variance explained. This is weak to moderate evidence that the Current Violence scale is unidimensional.

Local Independence The assumption of local independence was tested using the Chi-Squared Test of Independence in R with the parameter `simulate.p.value` set to `TRUE`. The n = 1,000 subset used for testing DIF from the transformed data (items changed for the COMPAS-R Core scales) was used. There was 8 pair of items with p-values greater than or equal to 0.05. These item pairs are listed in Table 49.

[4] n = 2,000 was used because n = 1,000 did not produce enough heterogeneity in the data among race to test for DIF: several subsamples with n = 1,000 were tested.

Table 49: Item pairs and their corresponding p-value for items that suggest independence from a subset of n = 1,000 from the final COMPAS-R CORE scale.

item1	item2	p_value
curr_assault	curr_homicide	0.402
curr_assault	curr_robbery	0.800
curr_assault	curr_weapons	0.912
curr_sexwithforce	curr_homicide	0.753
curr_sexwithforce	curr_robbery	1.000
curr_homicide	curr_robbery	0.195
curr_homicide	curr_domviol	0.132
curr_weapons	curr_domviol	0.170

Latent Monotonicity The latent monotonicity assumption is investigated by examining the observable property of the test data known as manifest monotonicity. We see in Table 50 that there are several violations (#vi).

Table 50: Manifest monotonicity results: used to assess latent monotonicity. The function `check.mononotinicity` was used from the R package *mokken*.

	ItemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
crime_cat	0.55	3	0	0.00	0.00	0.00	0.0000	0.00	0	0
curr_assault	0.27	3	1	0.33	0.09	0.09	0.0307	7.61	1	160
curr_sexwithforce	0.11	3	1	0.33	0.11	0.11	0.0364	13.45	1	204
curr_homicide	0.17	6	1	0.17	0.04	0.04	0.0075	5.72	1	93
curr_robbery	0.16	3	1	0.33	0.08	0.08	0.0253	8.82	1	164
curr_weapons	0.10	3	0	0.00	0.00	0.00	0.0000	0.00	0	0
curr_domviol	0.22	3	0	0.00	0.00	0.00	0.0000	0.00	0	0

7.5 Associates and Peers

The Associates and Peers scale is derived from the Criminal Associates and Peers scale in the standard COMPAS Core. In the development sample, there were 16,089 individuals with completed Criminal Associates and Peers scales.

Review of Theoretical and Substantive Validity

An involvement with anti-social friends and associates is one of the major risk factors for recidivism (Andrews and Bonta 2016).

Scale Items

The Associates and Peers scale is constructed using five items from the Relationships/Lifestyle domain of the COMPAS-R Core. These items and their short descriptions are shown in Table 51.

As stated in the original COMPAS Core psychometric study (Brennan, Dieterich, and Oliver 2007), the purpose of this scale is to evaluate a person's level of involvement with peers who have been arrested or incarcerated, take illegal drugs, or belong to gangs, including whether they, themselves, belong to a gang. A high score indicates a person associated with a network of delinquent friends.

Table 51: COMPAS-R Core Associates and Peers Scale Items. These items are in the interview portion of the assessment.

Item	Short Description (Response Categories)
friends_arrest_R	How many of your friends have ever been arrested? *(None=0; Few=1; Half=2; Most=3)*
friends_drugs_R	How many of your friends are taking illegal drugs regularly (more than a couple times a month)? *(None=0; Few=1; Half=2; Most=3)*
friends_gang_R	How many of your friends are gang members? *(None=0; Few=1; Half=2; Most=3)*
friends_jail_R	How many of your friends are/were in jail or prison? *(None=0; Few=1; Half=2; Most=3)*
gang_member_R	Are you now a gang member? *(No=0; Yes=1)*

Item changes from the Criminal Associates and Peers scale in the standard COMPAS Core

Two items from the Criminal Associates and Peers scale of the standard COMPAS Core are not present in the Associates and Peers scale of the COMPAS-R Core. These items are 1) the practitioner's opinion of whether the person is a gang member and 2) the interview item that asks the respondent whether they have ever been a gang member. This shortens the scale to five items.

This change increased the reliability of the scale from 0.748 to 0.768, and it is well within an acceptable range of reliability (Cortina 1993). Additional statistics for the scale are included in this section.

Factor structure in the norm data

Table 52: Principal component loadings.

	PC1	PC2	PC3
friends_arrest	-0.606	0.407	-0.136
friends_drugs	-0.549	-0.821	-0.156
friends_gang	-0.167	-0.076	0.968
friends_jail	-0.551	0.393	0.007
gang_member	-0.020	-0.016	0.140

Table 53: Eigenvalues and the proportion of the variance explained by each principal component.

	PC1	PC2	PC3
Eigenvalue	1.977	0.447	0.215
Standard deviation	1.406	0.669	0.464
Proportion of variance	0.707	0.160	0.077
Cumulative proportion	0.707	0.867	0.944

Principal components analysis of the Associates and Peers items in the COMPAS-R composite norm group is presented here. The loadings of the first three principal components are shown in Table 52. Table 53 shows the eigenvalues of the principal components and the proportion of the variance explained by each. As shown in the table, the first principal component has an eigenvalue of 1.977 and explains 70.7% of the variance in the data. Three principal components are required to explain at least 90% of the variation in the data.

Scale consistency and reliability in the norm data

Table 54: Correlation of items with scale score, and alpha reliability of the scale when each item is removed. Compare with alpha of 0.768 for all five items.

	Correlation with score	Alpha when item removed
friends_arrest	0.889	0.630
friends_drugs	0.808	0.705
friends_gang	0.580	0.762
friends_jail	0.881	0.632
gang_member	0.265	0.809

Cronbach's alpha measures a scale's internal consistency and reliability. Alpha for the Associates and Peers scale in the composite norm set is 0.768. Table 54 shows the correlation of each item with the overall scale score, and the value of alpha when the item is removed from the scale. Removing the item *gang_member* would increase alpha, but its current value is sufficiently high. Also, it is important to have information about a person's gang membership. Removing any of the other items would lower alpha.

Scale distributions in the norm groups

Table 55: Tabulation by population and gender with mean scores. Note: this table does not present the scale score from the single transgender person in the data.

Group	N	Mean
Jail Female	404	4.0
Prison Female	957	2.7
Probation Female	1885	2.0
Jail Male	1900	3.4
Prison Male	4867	2.5
Probation Male	6075	2.2
ALL	16088	2.5

Table 55 shows the mean Associates and Peers scale score for each population subgroup in the normative data. Women in Jail have generally higher scores compared to Men. The mean scores for Men and Women in Prison are about the same, as are the mean scores for Men and Women in Probation. As expected, the scores of people in the Probation setting are lowest, on average.

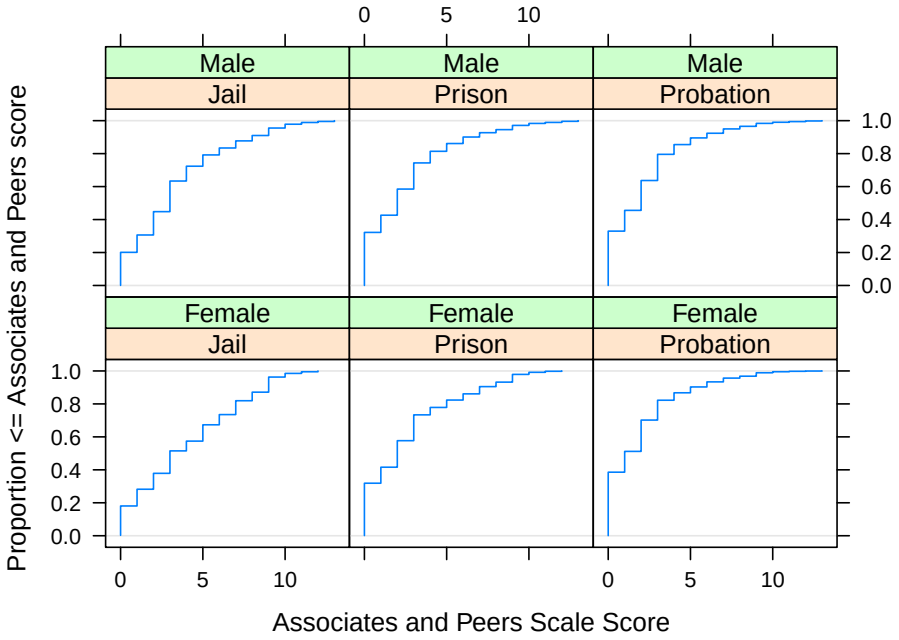


Figure 9: Cumulative distribution plots comparing subgroups in the norm data on the Associates and Peers scale

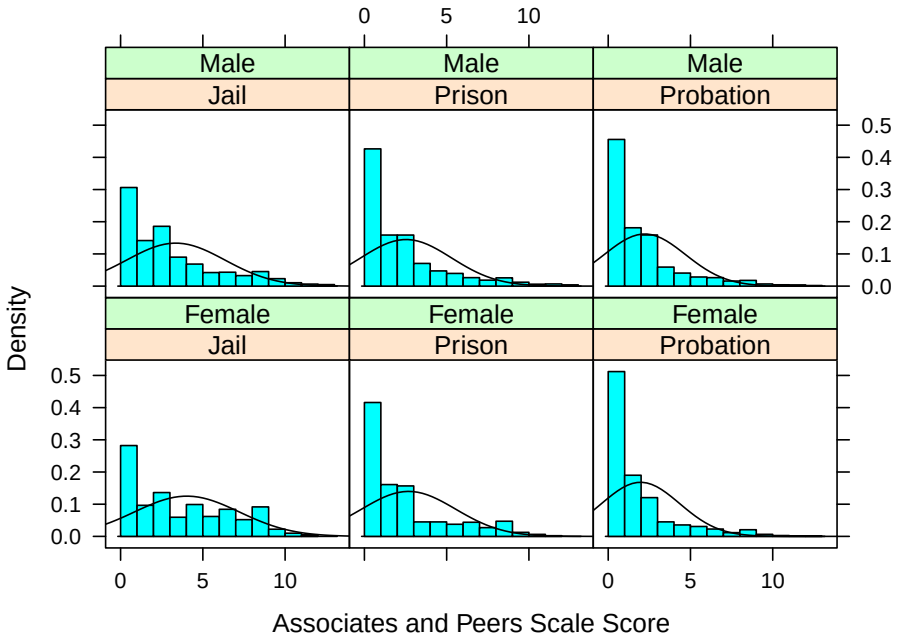


Figure 10: Histograms comparing subgroups in the norm data on the Associates and Peers scale

The cumulative distributions for the subgroups are shown in Figure 9. These plots enhance the information in Table 55, showing higher scores among those housed in Jail, especially Women.

The histograms for the subgroups are shown in Figure 10. These plots show the more even distribution of scores among those in Jail, compared to the scores for the prisoners and probationers, which show strong tendencies for very low scores.

Score ranges

Table 56 shows the score ranges for each of the subgroups in the norm data. GN is a subgroup that includes all the Men and Women. Since there are more Men in the data, the GN ranges more strongly reflect the ranges for Men than for Women. Most agencies choose to use ranges based on gender-specific norms; GN is not available in the software at this time.

The ranges are determined by splitting the scores of each subgroup into the lowest 40% of scores for Unlikely, the highest 30% of scores for Highly Probable, and the remaining 30% in between for Probable. This table of ranges reflects what was seen earlier in the score distributions by subgroup. Only the Jail groups stand out because of their Unlikely (40%) and Probable (30%) levels having higher scores in these levels than the other groups, particularly the Women.

Table 56: Score ranges for the Associates and Peers scale.

	Level		
	Unlikely	Probable	Highly Probable
GN Composite	0 – 1	2 – 3	4 – 13
GN Prison	0 – 1	2 – 3	4 – 13
GN Probation	0 – 1	2 – 3	4 – 13
GN Jail (Client-specific)	0 – 2	3 – 5	6 – 13
Male Composite	0 – 1	2 – 3	4 – 13
Male Prison	0 – 1	2 – 3	4 – 13
Male Probation	0 – 1	2 – 3	4 – 13
Male Jail (Client-specific)	0 – 2	3 – 4	5 – 13
Female Composite	0 – 1	2 – 3	4 – 13
Female Prison	0 – 1	2 – 3	4 – 13
Female Probation	0 – 1	2	3 – 13
Female Jail (Client-specific)	0 – 3	4 – 6	7 – 13

Mokken Scale Analysis (MSA): Checking the observable measurement properties of the Monotone Homogeneity Model

Item-pair scalability coefficient The final Associates and Peers scale variables are listed in Table 51. Table 57 displays the item-pair scalability coefficients for all Associates and Peers scale items. All values are positive with the smallest scalability coefficient (0.262) between *friends_jail and gang_member*. The largest scalability coefficient is 0.893 between *friends_arrest and friends_jail*.

Table 57: Item-pair scalability coefficients (H_{ij}). Higher H_{ij} result in stronger scales; $-\infty < H_{ij} \leq 1$. The monotone homogeneity (IRT) model assumption requires that $0 \leq H_{ij} \leq 1$.

	friends_arrest	se	friends_drugs	se	friends_gang	se	friends_jail	se	gang_member	se
friends_arrest			0.629	(0.007)	0.500	(0.012)	0.893	(0.004)	0.301	(0.023)
friends_drugs	0.629	(0.007)			0.499	(0.012)	0.605	(0.008)	0.331	(0.022)
friends_gang	0.500	(0.012)	0.499	(0.012)			0.502	(0.012)	0.470	(0.019)
friends_jail	0.893	(0.004)	0.605	(0.008)	0.502	(0.012)			0.262	(0.021)
gang_member	0.301	(0.023)	0.331	(0.022)	0.470	(0.019)	0.262	(0.021)		

* The smallest item-pair scalability coefficient is 0.262 and the largest is 0.893.

Item-scalability coefficient The item scalability coefficients are in Table 58; they are all above the lower bound of 0.3.

Table 58: Mokken Analysis: Item H values should be above 0.3

	Item H	se
friends_arrest	0.697	(0.005)
friends_drugs	0.587	(0.007)
friends_gang	0.498	(0.009)
friends_jail	0.689	(0.005)
gang_member	0.343	(0.018)

Test Scalability Coefficient The test scalability coefficient (H) for the Associates and Peers scale in its final form is 0.621 which is within the limits for the monotone homogeneity model. A scalability coefficient of $H = 0.621$ implies that the scale is strong. The standard error for the statistic is (0.005). The original data had a test scalability coefficient of 0.546 with a standard error of (0.006).

Differential Item Functioning (DIF) Differential Item Functioning (DIF) was tested on a subset of the data ($n = 1{,}000$) for race, after variable transformations; the *Other* category for race was not included since it only made up 7.5% of the final sample. Table 59 denotes the ethnicity proportions from the random subset of $n = 1{,}000$ from the transformed data.

The following items were found in the random subsample of $n = 1{,}000$ to have differential item functioning (DIF): *friends_gang and friends_jail*. In the original data, a random subsample of $n = 1{,}000$ was also selected; the following variables were flagged with DIF: *gang_obs, friends_gang, gang_member and prev_gang_mem*.

Table 59: Percent of Black, Hispanic, and White persons in a random sample of 1,000 cases drawn from the composite norm data; the data includes all variable transformations.

Ethnicity	Percent
Black	39
Hispanic	15
White	46

Non-Parametric Monotone Homogeneity (IRT) Model

The three assumptions for the monotone homogeneity model are the *Unidimensionality*, *Local Independence*, and *Latent Monotonicity*.

Unidimensionality The Unidimensionality assumption is tested using Principal Components Analysis (PCA). This was done in section 7.5. We see in Table 53 that the first eigenvalue (1.977) is much larger than the remaining two and over half (70.7%) of the variance is explained. This is strong evidence that the Associates and Peers scale is unidimensional.

Local Independence The assumption of local independence was tested using the Chi-Squared Test of Independence in R with the parameter `simulate.p.value` set to `TRUE`. The $n = 1{,}000$ subset used for testing DIF from the transformed data (items changed for the COMPAS-R Core scales) was used. All p-values were found to be less than 0.001, indicating evidence of local dependence.

Latent Monotonicity The latent monotonicity assumption is investigated by examining the observable property of the test data known as manifest monotonicity. We see in Table 60 that there are no violations (#vi).

Table 60: Manifest monotonicity results: used to assess latent monotonicity. The function `check.mononotinicity` was used from the R package *mokken*.

	ItemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
friends_arrest	0.70	18	0	0	0	0	0	0	0	0
friends_drugs	0.59	30	0	0	0	0	0	0	0	0
friends_gang	0.50	27	0	0	0	0	0	0	0	0
friends_jail	0.69	30	0	0	0	0	0	0	0	0
gang_member	0.34	15	0	0	0	0	0	0	0	0

7.6 Substance Use

The Substance Use scale is derived from the Substance Abuse scale in the standard COMPAS Core. In the development sample, there were 17,926 individuals with completed Substance Abuse scales.

Review of Theoretical and Substantive Validity

The misuse of alcohol and drugs is a major recidivism risk factor, among the Central Eight criminogenic needs (Andrews and Bonta 2016).

Scale Items

The Substance Use scale is constructed using six items from the Relationships/Lifestyle domain of the COMPAS-R Core. These items and their short descriptions are shown in Table 61.

The Substance Use scale is a general indicator of substance use problems, similar to a screening tool. A high score suggests the person may have a serious drug or alcohol problem and should be referred for a more detailed substance use test, such as the UNCOPE (Hoffmann et al. 2003). The items in this scale cover prior and current treatment for alcohol or drug problems, whether a person believes treatment would be beneficial, whether the person blames drugs or alcohol for their present problems, and whether the person was using alcohol or drugs at the time of arrest for the current offense.

The item labeled *op_arrest* in Table 61 is new. Although it does not contribute to the scale score, an agency can configure the scale set so that an affirmative response to this item moves the person into the next higher need level (unless their score is already Highly Probable), with the purpose of facilitating their access to substance use treatment.

Table 61: COMPAS-R Core Substance Use Scale Items

Item	Short Description (Response Categories)
ad_arrest_R	Were you using alcohol or drugs (including opioids) at the time of the current offense? *(No=0; Yes, alcohol only=1; Yes, drugs only=1; Yes, both=2)*
benefit_rx_ad_R	Do you think you would benefit from treatment for alcohol, or drugs, or both? *(No=0; Yes, alcohol only=1; Yes, drugs only=1; Yes, both=2)*
blame_drugs_R	Do you think your legal problems are due to alcohol or drugs? *(No=0; Yes=1)*
current_treatment_R	Are you currently in treatment for alcohol or drugs? *(No=0; Yes=1)*
ever_rx_ad_R	Have you ever been in formal treatment for alcohol or drugs, such as counseling, outpatient, inpatient, or residential? *(No=0; Yes, alcohol only=1; Yes, drugs only=1; Yes, both=2)*
op_arrest_R	Is the current offense opioid related (were opioids involved at the time of offense or arrest)? *(No; Yes; Not scored*)*

* This item is for data collection and is not used to score the Substance Use scale. However, during scale set configuration, agencies can specify that an affirmative response to this item be used to increase a person's Substance Use need level by default. The goal in making this configuration option available is to help ensure appropriate supervision for opioid-involved individuals.

Item changes from the Substance Abuse scale in the standard COMPAS Core

Two items are removed from the scale for the COMPAS-R: the DUI item from the Current Charges table, and the item that asks if the person used heroin, cocaine, crack or methamphetamines as a juvenile.

Additionally, three pairs of items (six total) were collapsed to form three single items. For example, the item that asks if the person thinks they would benefit from getting treatment for drugs and the item that asks if the person thinks they would benefit from getting treatment for alcohol, were combined into a single item asking if the person thinks they would benefit from getting treatment for alcohol or drugs, with responses adjusted accordingly to collect the same information as in the original scales.

These changes reduce the reliability of the scale from 0.76 to 0.714, but it is still within an acceptable range of reliability (Cortina 1993). Additional statistics for the scale are included in this section.

Factor structure in the norm data

Table 62: Principal component loadings.

	PC1	PC2	PC3	PC4	PC5
blame_drugs	-0.335	0.148	-0.104	0.921	0.085
current_treatment	-0.107	-0.002	0.073	0.061	-0.990
ad_arrest	-0.519	0.501	-0.605	-0.337	-0.010
ever_rx_ad	-0.583	-0.797	-0.105	-0.100	0.050
benefit_rx_ad	-0.516	0.301	0.779	-0.158	0.103

Table 63: Eigenvalues and the proportion of the variance explained by each principal component.

	PC1	PC2	PC3	PC4	PC5
Eigenvalue	1.082	0.406	0.295	0.136	0.084
Standard deviation	1.040	0.637	0.543	0.368	0.289
Proportion of variance	0.540	0.203	0.147	0.068	0.042
Cumulative proportion	0.540	0.743	0.890	0.958	1.000

Principal components analysis of the Substance Use items in the COMPAS-R composite norm group is presented here. The loadings of the principal components are shown in Table 62. Table 63 shows the eigenvalues of the principal components and the proportion of the variance explained by each. As shown in the table, the first principal component has an eigenvalue of 1.082 and explains 54% of the variance in the data. Four principal components are needed to explain at least 90% of the variation in the data.

Scale consistency and reliability in the norm data

Table 64: Correlation of items with scale score, and alpha reliability of the scale when each item is removed. Compare with alpha of 0.714 for all five items.

	Correlation with score	Alpha when item removed
blame_drugs	0.739	0.631
current_treatment	0.456	0.722
ad_arrest	0.749	0.641
ever_rx_ad	0.723	0.685
benefit_rx_ad	0.755	0.636

Cronbach's alpha measures a scale's internal consistency and reliability. Alpha for the Substance Use scale in the composite norm set is 0.714. Table 64 shows the correlation of each item with the overall scale score, and the value of alpha when the item is removed from the scale. Removing the item that asks of the person is currently in treatment increases alpha, but it is a useful item to have information about and is well-correlated with the scale score. Removing any of the other items would lower alpha substantially.

Scale distributions in the norm groups

Table 65: Tabulation by population and gender with mean scores. Note: this table does not present the scale score from the single transgender person in the data.

Group	N	Mean
Jail Female	405	2.9
Prison Female	1149	2.6
Probation Female	2167	2.1
Jail Male	1900	2.5
Prison Male	5572	2.3
Probation Male	6732	2.2
ALL	17925	2.3

Table 65 shows the mean Substance Use scale score for each population subgroup in the normative data. While all groups have mean scores close to the overall mean score of 2.3, Jail inmates and Women prisoners are higher than the mean, especially Jail inmates who are Women.

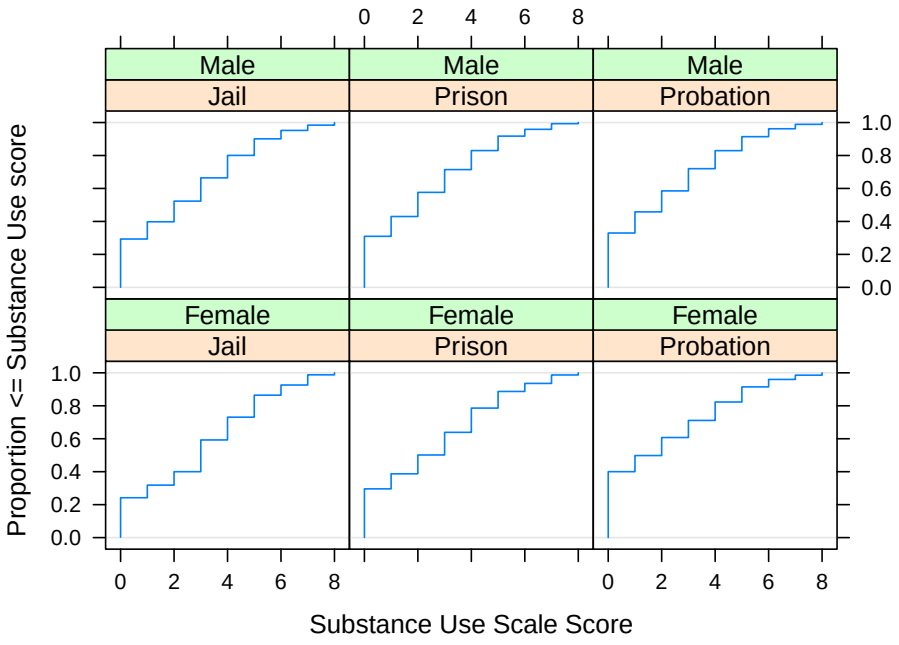


Figure 11: Cumulative distribution plots comparing subgroups in the norm data on the Substance Use scale

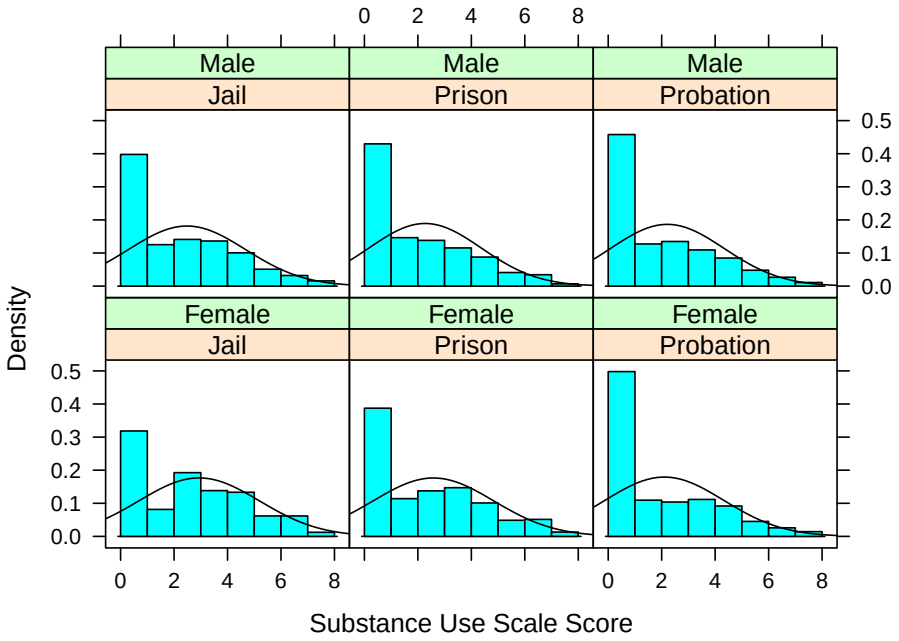


Figure 12: Histograms comparing subgroups in the norm data on the Substance Use scale

The cumulative distributions for the subgroups are shown in Figure 11 and corresponding histograms are shown in Figure 12. These plots enhance the information in Table 65. All groups have a large proportion of people scoring 0 on this scale, and score distributions above 0 are very similar.

Score ranges

Table 66 shows the score ranges for each of the subgroups in the norm data. GN is a subgroup that includes all the Men and Women. Since there are more Men in the data, the GN ranges more strongly reflect the ranges for Men than for Women. Most agencies choose to use ranges based on gender-specific norms; GN is not available in the software at this time.

The ranges are determined by splitting the scores of each subgroup into the lowest 20% of scores for Unlikely, the highest 60% of scores for Highly Probable, and the remaining 20% in between for Probable. This table of ranges reflects what was seen in the previous section for score distributions by subgroup. For example, the very large proportion of Women probationers who scored zero on the assessment are Unlikely to have a Substance Use problem, leaving those who scored 1 or higher to make up the Highly Probable range of scores.

Table 66: Score ranges for the Substance Use scale.

	Level		
	Unlikely	Probable	Highly Probable
GN Composite	0	1	2 – 8
GN Prison	0	1	2 – 8
GN Probation	0	1	2 – 8
GN Jail (Client-specific)	0	1 – 2	3 – 8
Male Composite	0	1	2 – 8
Male Prison	0	1	2 – 8
Male Probation	0	1	2 – 8
Male Jail (Client-specific)	0	1 – 2	3 – 8
Female Composite	0	1	2 – 8
Female Prison	0	1 – 2	3 – 8
Female Probation	0		1 – 8
Female Jail (Client-specific)	0	1 – 2	3 – 8

Mokken Scale Analysis (MSA): Checking the observable measurement properties of the Monotone Homogeneity Model

Item-pair scalability coefficient The final Substance Use scale variables are listed in Table 61. Table 67 displays the item-pair scalability coefficients for all Substance Use scale items. All values are positive with the smallest scalability coefficient (0.294) between ***current_treatment and ad_arrest***. The largest scalability coefficient is 0.651 between ***blame_drugs and benefit_rx_ad***.

Table 67: Item-pair scalability coefficients (H_{ij}). Higher H_{ij} result in stronger scales; $-\infty < H_{ij} \leq 1$. The monotone homogeneity (IRT) model assumption requires that $0 \leq H_{ij} \leq 1$.

	blame_drugs	se	current_treatment	se	ever_rx_ad	se	benefit_rx_ad	se	ad_arrest	se
blame_drugs			0.645	(0.016)	0.450	(0.008)	0.651	(0.009)	0.621	(0.008)
current_treatment	0.645	(0.016)			0.415	(0.012)	0.403	(0.010)	0.294	(0.011)
ever_rx_ad	0.450	(0.008)	0.415	(0.012)			0.415	(0.009)	0.349	(0.008)
benefit_rx_ad	0.651	(0.009)	0.403	(0.010)	0.415	(0.009)			0.477	(0.008)
ad_arrest	0.621	(0.008)	0.294	(0.011)	0.349	(0.008)	0.477	(0.008)		

* The smallest item-pair scalability coefficient is 0.294 and the largest is 0.651.

Item-scalability coefficient The item scalability coefficients are in Table 68; they are all above the lower bound of 0.3.

Table 68: Mokken Analysis: Item H values should be above 0.3

	Item H	se
blame_drugs	0.569	(0.006)
current_treatment	0.401	(0.008)
ever_rx_ad	0.400	(0.006)
benefit_rx_ad	0.477	(0.005)
ad_arrest	0.442	(0.005)

Test Scalability Coefficient The test scalability coefficient (H) for the Substance Use scale in its final form is 0.456 which is within the limits for the monotone homogeneity model. A scalability coefficient of $H = 0.456$ implies that the scale is moderate. The standard error for the statistic is (0.005). The original data had a test scalability coefficient of 0.359 with a standard error of (0.004).

Differential Item Functioning (DIF) Differential Item Functioning (DIF) was tested on a subset of the data (n = 1,000) for race, after variable transformations; the *Other* category for race was not included since it only made up 7.2% of the final sample. Table 69 denotes the ethnicity proportions from the random subset of n = 1,000 from the transformed data.

The following items were found to have differential item functioning (DIF) in the n = 1,000 subsample: ***blame_drugs, ever_rx_ad and ad_arrest***. In the original data, a subsample of n = 1,000 was also selected; the following variables were flagged with DIF: ***curr_dui, want_rx_a, blame_drugs and juv_drugs***. The variables ***curr_dui*** and ***juv_drugs*** were not included in the final Substance Use scale.

Non-Parametric Monotone Homogeneity (IRT) Model

The three assumptions for the monotone homogeneity model are the *Unidimensionality*, *Local Independence*, and *Latent Monotonicity*.

Table 69: Percent of Black, Hispanic, and White persons in a random sample of 1,000 cases drawn from the composite norm data; the data includes all variable transformations.

Ethnicity	Percent
Black	44
Hispanic	12
White	45

Unidimensionality The Unidimensionality assumption is tested using Principal Components Analysis (PCA). This was done in section 7.6. We see in Table 63 that the first eigenvalue is larger than the remaining four (1.146) and over half (54.0%) of the variance is explained. This is moderate to strong evidence that the Substance Use scale is unidimensional.

Local Independence The assumption of local independence was tested using the Chi-Squared Test of Independence in R with the parameter `simulate.p.value` set to `TRUE`. All p-values were found to be less than 0.001, indicating evidence of local dependence.

Latent Monotonicity The latent monotonicity assumption is investigated by examining the observable property of the test data known as manifest monotonicity. We see in Table 70 that there are no violations (#vi).

Table 70: Manifest monotonicity results: used to assess latent monotonicity. The function `check.mononotinicity` was used from the R package *mokken*.

	ItemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
blame_drugs	0.57	10	0	0	0	0	0	0	0	0
current_treatment	0.40	15	0	0	0	0	0	0	0	0
ever_rx_ad	0.40	20	0	0	0	0	0	0	0	0
benefit_rx_ad	0.48	20	0	0	0	0	0	0	0	0
ad_arrest	0.44	20	0	0	0	0	0	0	0	0

7.7 Society and Routines

The Society and Routines scale is derived from Criminal Opportunity scale in the standard COMPAS Core. In the development sample, there were 12,956 individuals with completed Criminal Opportunity scales. In the analysis, the decision was made to remove the answer choice "Unsure" from an item. This reduced the number of individuals in the sample to 11,851.

Review of Theoretical and Substantive Validity

This higher order scale was originally developed to assess the concept of criminal opportunity, reflected in criminological theories that emphasize routine activities and social roles (spouse, parent, employee, etc.) that reinforce pro-social attitudes, social bonds and associated social controls.

From the original COMPAS Psychometric study (Brennan, Dieterich, and Oliver 2007):

> The theoretical background to this scale includes routine activities theory that emphasizes the importance of immediate local daily activities that place a person in high risk or high opportunity situations (Cohen and Felson 1979). The second theoretical theme contributing to this scale is early social control theory (Hirschi 1969) which emphasizes the importance of social bonds as inhibitors or constraints to crime.

Scale Items

The Society and Routines scale is constructed using twelve items from the Relationships/Lifestyle domain of the COMPAS-R Core. These items and their short descriptions are shown in Table 71.

The Society and Routines scale uses items that represent a combination of spending time in high crime situations, affiliating with persons who are likely to engage in illegal activities, and an absence of pro-social or constructive activities (e.g., working, spending time with family).

Item changes from the Criminal Opportunity scale in the standard COMPAS Core to the Society and Routines scale in the COMPAS-R Core

The item on the standard COMPAS Core that asks if there is a telephone at the person's residence was removed. Another item that is calculated behind-the-scenes, "age group," was also removed. Both of these items had very low correlations with the overall scale score and their removal resulted in a modest increase in scale reliability, from 0.636 to 0.640.

Responses on some retained items were changed to facilitate increased accuracy:

1. The responses to the item that asks how often the person has felt recently that they had nothing to do in their free time were changed from: Never, Several times a month, Several times a week, and Daily, to simply: Never, Occasionally, and Daily.

2. The "Unsure" response choice was removed from the item asking if the person had recently often become bored with their usual activities. There were 1,105 individuals in the composite norm set who responded that they were unsure about this question, reducing the number in the sample from 12,956 to 11,851.

These changes reduced alpha to 0.622, but it is not much lower than it started out (0.636).

Additional statistics for the scale are included in this section.

Table 71: COMPAS-R Core Society and Routines Scale Items

Item	Short Description (Response Categories)
citizens_weapons_R	Do neighbors feel the need to carry a weapon for protection? *(No=0; Yes=1)*
friends_drugs_R	How many of your friends are taking illegal drugs regularly (more than a couple times a month)? *(None=0; Few=1; Half=2; Most=3)*
haveempschool_R	Do you have a verifiable employer or school? *(No=1; Yes=0)*
job_last_year_R	How much have you worked or been enrolled in school within the last 12 months? *(12 months full time=0; 12 months part time=1; More than 6 months (up to, but not including a year), either part time or full time=2; Between 0 and 6 months, either part time or full time=3)*
job_R	Do you currently have a job? *(No=1; Yes=0)*
l_drift_R	Do you have a regular living situation (an address where you usually stay and can be reached)? *(No=1; Yes=0)*
local_crime_R	Is there much crime where you generally live? *(No=0; Yes=1)*
local_gangs_R	Are there gangs in your neighborhood? *(No=0; Yes=1)*
need_training_R	Right now, do you feel you need more training in a new job or career skill? *(No=0; Yes=1)*
nothing_R	In the past few months, how often have you felt you had nothing to do in your free time? *(Never=0; Occasionally=1; Daily=2)*
res_moves_R	How often have you moved in last 12 months? *(None=0; Once=1; Twice=2; Three times=3; Four times=4; Five or more times=5)*
restless_R	In the past few months, did you often become bored with your usual activities? *(No=0; Yes=1)*

Factor structure in the norm data

Table 72: Principal component loadings.

	PC1	PC2	PC3	PC4	PC5	PC6
citizens_weapons	0.027	-0.038	0.074	-0.015	-0.072	0.338
friends_drugs	0.174	-0.369	0.839	0.312	-0.002	-0.171
haveempschool	0.198	0.076	0.000	-0.089	-0.656	-0.151
job	0.208	0.085	-0.008	-0.103	-0.638	-0.154
job_last_year	0.864	0.383	-0.052	0.137	0.289	0.037
l_drift	0.042	-0.069	-0.023	-0.029	-0.030	-0.004
local_crime	0.036	-0.043	0.079	-0.016	-0.112	0.452
local_gangs	0.028	-0.040	0.072	-0.010	-0.078	0.357
need_training	0.064	-0.045	0.105	-0.080	-0.125	0.680
nothing	0.135	-0.095	0.216	-0.857	0.190	-0.115
res_moves	0.331	-0.824	-0.451	0.041	0.006	-0.017
restless	0.068	-0.068	0.120	-0.347	0.044	0.026

Table 73: Eigenvalues and the proportion of the variance explained by each principal component.

	PC1	PC2	PC3	PC4	PC5	PC6
Eigenvalue	2.092	1.397	0.874	0.404	0.267	0.253
Standard deviation	1.446	1.182	0.935	0.636	0.517	0.503
Proportion of variance	0.358	0.239	0.149	0.069	0.046	0.043
Cumulative proportion	0.358	0.596	0.746	0.815	0.860	0.904

Principal components analysis of the Society and Routines items in the COMPAS-R composite norm group is presented here. The loadings of the first six principal components are shown in Table 72. Table 73 shows the eigenvalues of the first six principal components and the proportion of the variance explained by each. As shown in the table, the first principal component has an eigenvalue of 2.092 and explains 35.8% of the variance in the data. Six principal components are required to explain at least 90% of the variation in the data.

Scale consistency and reliability in the norm data

Table 74: Correlation of items with scale score, and alpha reliability of the scale when each item is removed. Compare with alpha of 0.622 for all twelve items.

	Correlation with score	Alpha when item removed
citizens_weapons	0.351	0.609
friends_drugs	0.524	0.599
haveempschool	0.545	0.580
job	0.550	0.579
job_last_year	0.645	0.606
l_drift	0.336	0.611
local_crime	0.372	0.606
local_gangs	0.351	0.609
need_training	0.358	0.609
nothing	0.499	0.587
res_moves	0.534	0.623
restless	0.446	0.597

Cronbach's alpha measures a scale's internal consistency and reliability. Alpha for the Society and Routines scale in the composite norm set is 0.622. Table 74 shows the correlation of each item with the overall scale score, and the value of alpha when the item is removed from the scale. Removing the number of residential moves in the previous 12 months item affects alpha minimally, but it is very well-correlated with the scale score. Removing any of the other items would lower alpha.

Scale distributions in the norm groups

Table 75: Tabulation by population and gender with mean scores.

Group	N	Mean
Jail Female	365	7.6
Prison Female	593	6.8
Probation Female	1350	6.1
Jail Male	1722	5.3
Prison Male	3472	5.9
Probation Male	4349	5.5
ALL	11851	5.8

Table 75 shows the mean Society and Routines scale score for each population subgroup in the normative data. The mean score for Women is higher compared to Men, regardless of corrections setting; the mean score for Men is between 5 and 6. Among the scores for Women, the mean score of those on probation is closest to the mean scores for Men.

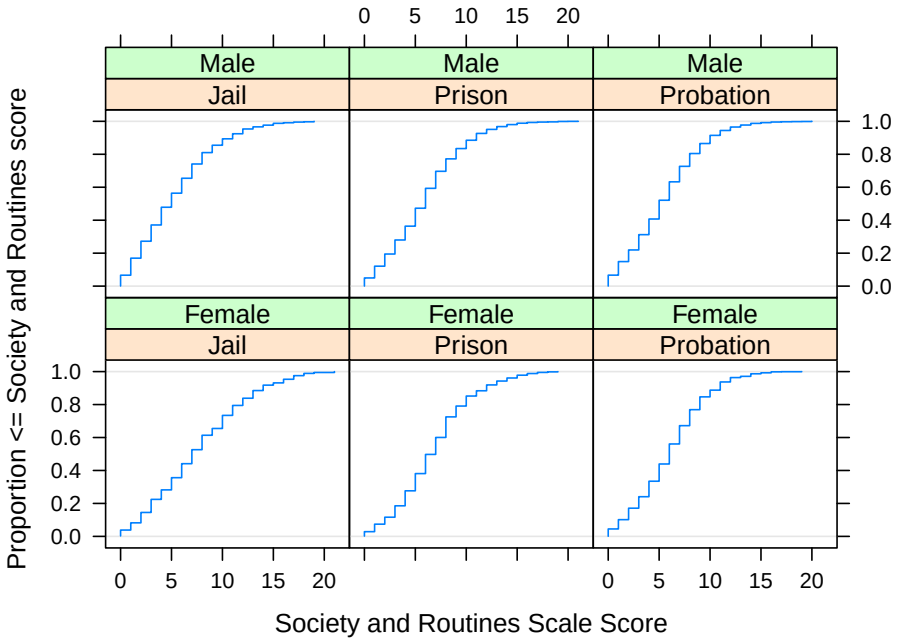


Figure 13: Cumulative distribution plots comparing subgroups in the norm data on the Society and Routines scale

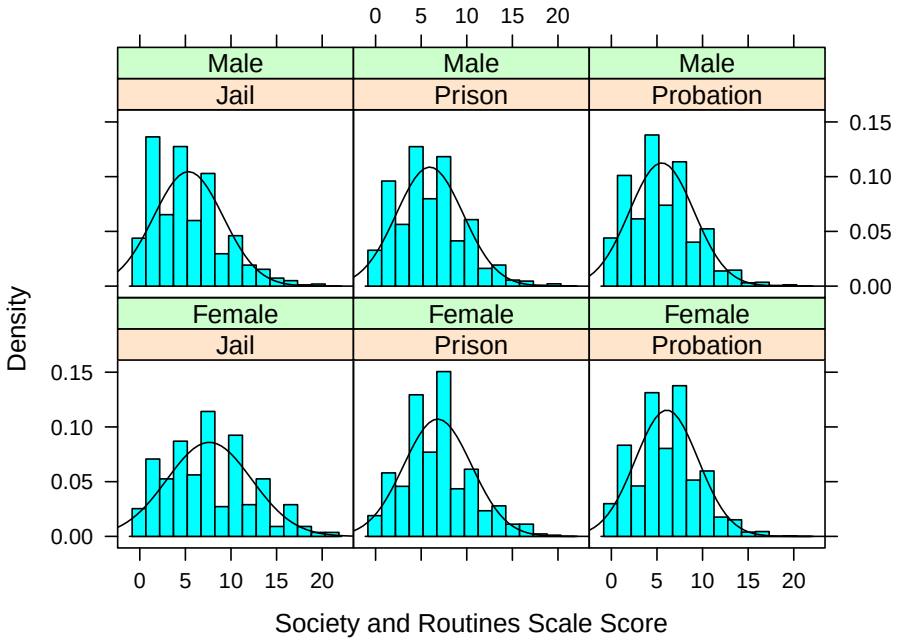


Figure 14: Histograms comparing subgroups in the norm data on the Society and Routines scale

The cumulative distributions for the subgroups are shown in Figure 13 and histograms are shown in Figure 14. These plots enhance the information in Table 75. The plots are all very similar, with

Table 76: Score ranges for the Society and Routines scale

	Level		
	Unlikely	Probable	Highly Probable
GN Composite	0 – 6	7	8 – 21
GN Prison	0 – 6	7 – 8	9 – 21
GN Probation	0 – 5	6 – 7	8 – 21
GN Jail (Client-specific)	0 – 5	6 – 7	8 – 21
Male Composite	0 – 5	6 – 7	8 – 21
Male Prison	0 – 6	7 – 8	9 – 21
Male Probation	0 – 5	6 – 7	8 – 21
Male Jail (Client-specific)	0 – 5	6 – 7	8 – 21
Female Composite	0 – 6	7 – 8	9 – 21
Female Prison	0 – 7	8	9 – 21
Female Probation	0 – 6	7 – 8	9 – 21
Female Jail (Client-specific)	0 – 7	8 – 10	11 – 21

the exception of Women in Jail, with scores that are less strongly right-skewed, explaining, in part, their higher mean score on this scale.

Score ranges

Table 76 displays the score ranges for each of the subgroups in the norm data. GN is a subgroup that includes all the Men and Women. Since there are more Men in the data, the GN ranges more strongly reflect the ranges for Men than for Women. Most agencies choose to use ranges based on gender-specific norms; GN is not available in the software at this time.

The ranges are determined by splitting the scores of each subgroup into the lowest 50% of scores for Unlikely, the highest 30% of scores for Highly Probable, and the remaining 20% in between for Probable.

Mokken Scale Analysis (MSA): Checking the observable measurement properties of the Monotone Homogeneity Model

Item-pair scalability coefficient The final Society and Routines scale variables are listed in Table 71. Table 77 displays the item-pair scalability coefficients for all Society and Routines scale items. All values are all positive. The smallest scalability coefficient (0.052) is between *job and friends_drugs*. The largest scalability coefficient is 0.847 between *job and haveempschool*.

11,851

Table 77: Item-pair scalability coefficients (H_{ij}). Higher H_{ij} result in stronger scales; $-\infty < H_{ij} \leq 1$. The monotone homogeneity (IRT) model assumption requires that $0 \leq H_{ij} \leq 1$.

	res_moves	l_drift	job	job_last_year	friends_drugs	local_crime	local_gangs	citizens_weapons	nothing	restless	need_training	haveempschool
res_moves		0.426	0.110	0.193	0.210	0.107	0.110	0.100	0.153	0.183	0.125	0.113
l_drift	0.426		0.294	0.342	0.151	0.108	0.064	0.078	0.180	0.206	0.150	0.291
job	0.110	0.294		0.685	0.052	0.102	0.080	0.108	0.201	0.200	0.093	0.847
job_last_year	0.193	0.342	0.685		0.111	0.163	0.155	0.160	0.260	0.242	0.135	0.708
friends_drugs	0.210	0.151	0.052	0.111		0.223	0.238	0.249	0.271	0.309	0.241	0.075
local_crime	0.107	0.108	0.102	0.163	0.223		0.652	0.645	0.132	0.124	0.286	0.125
local_gangs	0.110	0.064	0.080	0.155	0.238	0.652		0.509	0.133	0.146	0.291	0.104
citizens_weapons	0.100	0.078	0.108	0.160	0.249	0.645	0.509		0.145	0.173	0.320	0.080
nothing	0.153	0.180	0.201	0.260	0.271	0.132	0.133	0.145		0.648	0.168	0.170
restless	0.183	0.206	0.200	0.242	0.309	0.124	0.146	0.173	0.648		0.269	0.173
need_training	0.125	0.150	0.093	0.135	0.241	0.286	0.291	0.320	0.168	0.269		0.100
haveempschool	0.113	0.291	0.847	0.708	0.075	0.125	0.104	0.080	0.170	0.173	0.100	

* The standard errors (se) have been removed to allow all item names to fit in the table. The standard errors range from 0.007 to 0.03. The smallest item-pair scalability coefficient is 0.052 and the largest is 0.847.

Item-scalability coefficient The item scalability coefficients are in Table 78; they are not all above the lower bound of 0.3.

Table 78: Mokken Analysis: Item H values should be above 0.3

	Item H	se
res_moves	0.168	(0.007)
l_drift	0.238	(0.009)
job	0.305	(0.006)
job_last_year	0.288	(0.007)
friends_drugs	0.188	(0.006)
local_crime	0.212	(0.007)
local_gangs	0.214	(0.008)
citizens_weapons	0.219	(0.008)
nothing	0.225	(0.006)
restless	0.260	(0.007)
need_training	0.168	(0.006)
haveempschool	0.303	(0.006)

Test Scalability Coefficient The test scalability coefficient (H) for the Society and Routines scale in its final form is 0.228 which is within the limits for the monotone homogeneity model. A scalability coefficient of $H = 0.228$ implies that the scale is very weak. The standard error for the statistic is (0.004). The original data had a test scalability coefficient of 0.197 with a standard error of (0.003).

The Society and Routines scale is actually a higher order scale: multidimensional IRT was not

Table 79: Item scale determinations. Both methods conclude that there are actually three scales; however, the two methods do not agree in terms of which items belong together.

	Method 1	Method 2
res_moves	0	3
l_drift	1	3
job	1	1
job_last_year	1	1
friends_drugs	0	0
local_crime	2	1
local_gangs	2	1
citizens_weapons	2	1
nothing	3	2
restless	3	2
need_training	0	0
haveempschool	1	1

Table 80: Percent of Black, Hispanic, and White persons in a random sample of 1,000 cases drawn from the composite norm data; the data includes all variable transformations.

Ethnicity	Percent
Black	42
Hispanic	18
White	40

performed. Since the Society and Routines scale does not meet any Mokken scale assumptions, we do not test for any of the assumptions of the Monotone Homogeneity Model. Differential item functioning is still tested below. Table 79 displays the output from the R function `aisp` form the **mokken** package. It uses two algorithms (*Method 1* and *Method 2*) to partition items into Mokken scales. A zero (0) indicates that the item was not grouped/selected as a possible candidate for a mokken scale when combined with other items. Both methods conclude that there are actually three scales. The difference in algorithmic calculations separates the items differently into the three scales. For instance, **Method 1** groups *l_drift, job, job_last_year and haveempschool* as the first mokken scale, but **Method 2** groups *job, job_last_year, local_crime, local_gangs, citizens_weapons and haveempschool* together instead as the first Mokken scale found within the items.

Differential Item Functioning (DIF) Differential Item Functioning (DIF) was tested on a subset of the data (n = 1,000) for race, after variable transformations; the *Other* category for race was not included since it only made up 7.9% of the final sample. Table 80 denotes the ethnicity proportions from the random subset of n = 1,000 from the transformed data.

The following items were found to have differential item functioning (DIF) in the n = 1,000 subsample: *local_crime, local_gangs, citizens_weapons, restless and need_training*. In the original data, a subsample of n = 1,000 was also selected; the following variables were flagged with DIF: *res_moves, res_phone, local_gangs, citizens_weapons, nothing and haveempschool*.

7.8 Leisure and Recreation

The Leisure and Recreation scale is derived from the Leisure and Recreation scale in the standard COMPAS Core. In the development sample, there were 14,568 individuals with completed Leisure and Recreation scales. In the analysis, the decision was made to remove the answer choice "Unsure" from two items. This reduced the number of individuals in the sample to 12,761.

Review of Theoretical and Substantive Validity

Leisure/Recreation is included among the Central Eight Risk Needs Factors enumerated by Andrews and Bonta (2016). Theories of crime that support this association are Social Control theory (Hirschi 1969), General Theory of Crime (Gottfredson and Hirschi 1990), and Routine Activities theory (Osgood et al. 1996).

Scale Items

The Leisure and Recreation scale is constructed using four items from the Relationships/Lifestyle domain of the COMPAS-R Core. These items and their short descriptions are shown in Table 81.

This short scales assesses degree of boredom, restlessness, and lack of interest in current activities, reflecting the psychological dimension of the Leisure/Recreation need, rather than the amount of constructive opportunities available to the respondent. See Brennan, Dieterich, and Oliver (2007) for more details.

Table 81: COMPAS-R Core Leisure and Recreation Scale Items. Respondents are instructed to think of their leisure time in the past few months and, keeping in mind there are no wrong answers, indicate how often did they have the following feelings.

Item	Short Description (Response Categories)
bored_R	Thinking of your leisure time in the past few (3-6) months, how often did you feel bored? *(Never=0; Occasionally=1; Daily=2)*
do_boring_or_dull_R	In the past few months, did you feel the things you did were boring and dull? *(No=0; Yes=1)*
nothing_R	In the past few months, how often have you felt you had nothing to do in your free time? *(Never=0; Occasionally=1; Daily=2)*
restless_R	In the past few months, did you often become bored with your usual activities? *(No=0; Yes=1)*

Item changes from the Leisure and Recreation scale in the standard COMPAS Core

The item from the Leisure and Recreation scale in the standard COMPAS Core that asks whether it is difficult for the person to keep their mind on one thing was not retained.

Other changes to this scale were:

- the response option of "Unsure" was removed from *do_boring_or_dull* and *restless*.

- the responses of *Several times/month* and *Several times/week* were combined into *Occasionally*.

These changes reduced the reliability of the scale slightly, from 0.818 to 0.811, but it is still well-within an acceptable range of reliability (Cortina 1993). Removal of "Unsure" as an answer choice also resulted in a reduction of cases: there were 1,807 individuals in the composite norm set who responded that they were unsure about either *do_boring_or_dull* or *restless*, reducing the number in the sample from 14,568 to 12,761.

Additional statistics for the scale are included in this section.

Factor structure in the norm data

Table 82: Principal component loadings.

	PC1	PC2	PC3
bored	-0.683	-0.423	-0.595
do_boring_or_dull	-0.264	0.609	-0.128
nothing	-0.621	-0.091	0.778
restless	-0.278	0.665	-0.155

Table 83: Eigenvalues and the proportion of the variance explained by each principal component.

	PC1	PC2	PC3
Eigenvalue	0.862	0.165	0.122
Standard deviation	0.928	0.406	0.350
Proportion of variance	0.724	0.139	0.103
Cumulative proportion	0.724	0.863	0.966

Principal components analysis of the Leisure and Recreation items in the COMPAS-R composite norm group is presented here. The loadings of the first three principal components are shown in Table 82. Table 83 shows the eigenvalues of the principal components and the proportion of the variance explained by each. As shown in the table, the first principal component has an eigenvalue of 0.862 and explains 72.4% of the variance in the data. Three principal components are required to explain at least 90% of the variation in the data.

Scale consistency and reliability in the norm data

Table 84: Correlation of items with scale score, and alpha reliability of the scale when each item is removed. Compare with alpha of 0.811 for all four items.

	Correlation with score	Alpha when item removed
bored	0.867	0.749
do_boring_or_dull	0.752	0.782
nothing	0.869	0.722
restless	0.753	0.780

Cronbach's alpha measures a scale's internal consistency and reliability. Alpha for the Leisure and Recreation scale in the composite norm set is 0.811. Table 84 shows the correlation of each item with the overall scale score, and the value of alpha when the item is removed from the scale. Removing any of the remaining items would lower alpha.

Scale distributions in the norm groups

Table 85: Tabulation by population and gender with mean scores.

Group	N	Mean
Jail Female	339	2.4
Prison Female	800	1.6
Probation Female	1371	1.5
Jail Male	1622	1.8
Prison Male	4213	1.3
Probation Male	4416	1.4
ALL	12761	1.5

Table 85 shows the mean Leisure and Recreation scale score for each population subgroup in the normative data. Within each gender, those in Jail have the highest mean score on this scale, and within each corrections subgroup, Women have a higher mean score.

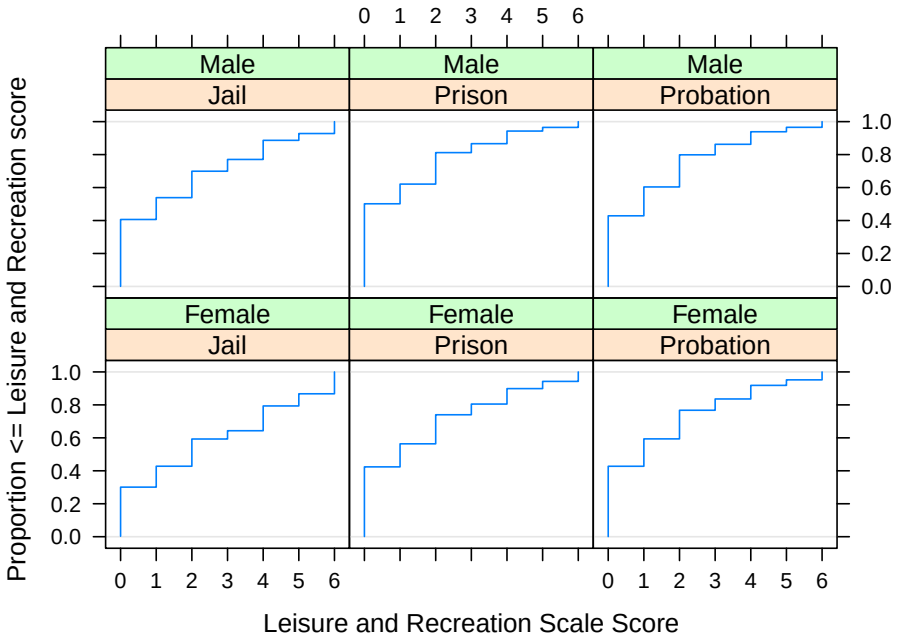


Figure 15: Cumulative distribution plots comparing subgroups in the norm data on the Leisure and Recreation scale

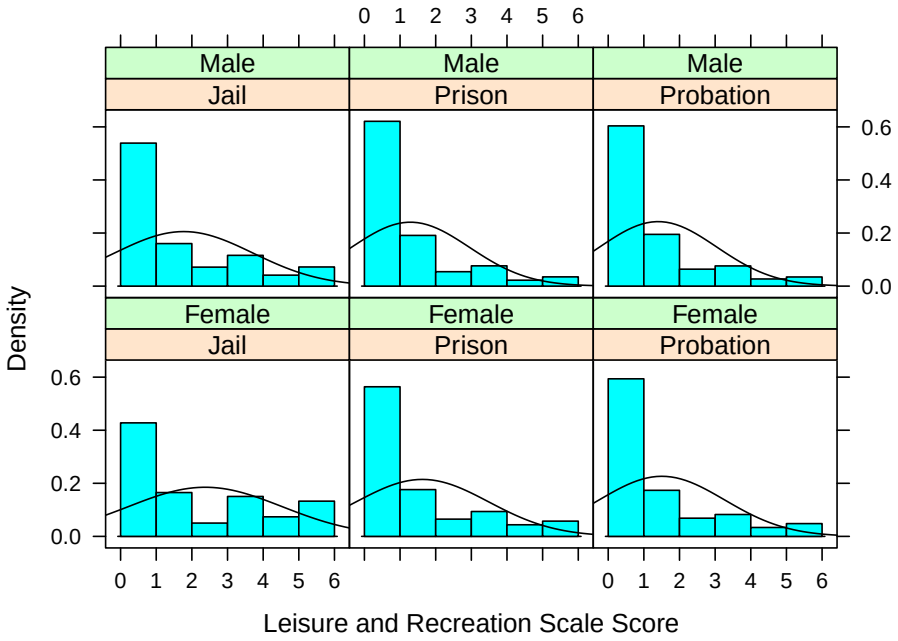


Figure 16: Histograms comparing subgroups in the norm data on the Leisure and Recreation scale

The cumulative distributions for the subgroups are shown in Figure 15 and histograms for the subgroups are shown in Figure 16. These plots enhance the information in Table 85. For example, Women appear to score higher than the Men in their respective corrections settings, with Women in Jail scoring highest. All subgroups have a substantial proportion scoring 0 on this scale.

Score ranges

Table 86 displays the score ranges for each of the subgroups in the norm data. GN is a subgroup that includes all the Men and Women. Since there are more Men in the data, the GN ranges more strongly reflect the ranges for Men than for Women. Most agencies choose to use ranges based on gender-specific norms; GN is not available in the software at this time.

The ranges are determined by splitting the scores of each subgroup into the lowest 50% of scores for Unlikely, the highest 30% of scores for Highly Probable, and the remaining 20% in between for Probable.

Table 86: Score ranges for the Leisure and Recreation scale.

	Level		
	Unlikely	Probable	Highly Probable
GN Composite	0 – 1	2	3 – 6
GN Prison	0 – 1	2	3 – 6
GN Probation	0 – 1	2	3 – 6
GN Jail (Client-specific)	0 – 1	2 – 3	4 – 6
Male Composite	0 – 1	2	3 – 6
Male Prison	0	1 – 2	3 – 6
Male Probation	0 – 1	2	3 – 6
Male Jail (Client-specific)	0 – 1	2 – 3	4 – 6
Female Composite	0 – 1	2	3 – 6
Female Prison	0 – 1	2	3 – 6
Female Probation	0 – 1	2	3 – 6
Female Jail (Client-specific)	0 – 2	3 – 4	5 – 6

Mokken Scale Analysis (MSA): Checking the observable measurement properties of the Monotone Homogeneity Model

Item-pair scalability coefficient The final Leisure and Recreation scale variables are listed in Table 81. Table 87 displays the item-pair scalability coefficients for all Leisure and Recreation scale items. All values are positive with the smallest scalability coefficient (0.654) between ***nothing and restless***. The largest scalability coefficient is 0.855 between ***bored and nothing***.

Table 87: Item-pair scalability coefficients (H_{ij}). Higher H_{ij} result in stronger scales; $-\infty < H_{ij} \leq 1$. The monotone homogeneity (IRT) model assumption requires that $0 \leq H_{ij} \leq 1$.

	bored	se	do_boring_or_dull	se	nothing	se	restless	se
bored			0.656	(0.011)	0.855	(0.006)	0.657	(0.010)
do_boring_or_dull	0.656	(0.011)			0.678	(0.011)	0.797	(0.009)
nothing	0.855	(0.006)	0.678	(0.011)			0.654	(0.011)
restless	0.657	(0.010)	0.797	(0.009)	0.654	(0.011)		

* The smallest item-pair scalability coefficient is 0.654 and the largest is 0.855.

Item-scalability coefficient The item scalability coefficients are in Table 88; they are all above the lower bound of 0.3.

Table 88: Mokken Analysis: Item H values should be above 0.3

	Item H	se
bored	0.754	(0.006)
do_boring_or_dull	0.703	(0.008)
nothing	0.761	(0.006)
restless	0.693	(0.008)

Test Scalability Coefficient The test scalability coefficient (H) for the Leisure and Recreation scale in its final form is 0.733 which is within the limits for the monotone homogeneity model. A scalability coefficient of $H = 0.733$ implies that the scale is strong. The standard error for the statistic is (0.006). The original data had a test scalability coefficient of 0.542 with a standard error of (0.005).

Differential Item Functioning (DIF) Differential Item Functioning (DIF) was tested on a subset of the data (n = 1,000) for race, after variable transformations; the *Other* category for race was not included since it only made up 7% of the final sample. Table 89 denotes the ethnicity proportions from the random subset of n = 1,000 from the transformed data.

No items were found to have differential item functioning (DIF) for the final Leisure scale subsample (n = 1,000). In the original data, the following variables were flagged with DIF: ***bored and scattered***. The item ***scattered*** was removed from the COMPAS-R Core scale Leisure and Recreation; it is still an item in the COMPAS Core scale of the same name. The item ***bored*** was not found to have DIF in the n = 1,000 subsample that was used above.

Table 89: Percent of Black, Hispanic, and White persons in a random sample of 1,000 cases drawn from the composite norm data; the data includes all variable transformations.

Ethnicity	Percent
Black	42
Hispanic	14
White	44

Non-Parametric Monotone Homogeneity (IRT) Model

The three assumptions for the monotone homogeneity model are the *Unidimensionality*, *Local Independence*, and *Latent Monotonicity*.

Unidimensionality The Unidimensionality assumption is tested using Principal Components Analysis (PCA). This was done in section 7.8. We see in Table 82 that the first eigenvalue (0.862) is larger than the remaining eigenvalues and over 70 percent (72.4%) of the variance is explained. This is strong evidence that the Leisure and Recreation scale is unidimensional.

Local Independence The assumption of local independence was tested using the Chi-Squared Test of Independence in R with the parameter `simulate.p.value` set to `TRUE`. All p-values were found to be less than 0.001, indicating evidence of local dependence.

Latent Monotonicity The latent monotonicity assumption is investigated by examining the observable property of the test data known as manifest monotonicity. We see in Table 90 that there are no violations (#vi).

Table 90: Manifest monotonicity results: used to assess latent monotonicity. The function `check.mononotinicity` was used from the R package *mokken*.

	ItemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
bored	0.75	6	0	0	0	0	0	0	0	0
do_boring_or_dull	0.70	6	0	0	0	0	0	0	0	0
nothing	0.76	6	0	0	0	0	0	0	0	0
restless	0.69	6	0	0	0	0	0	0	0	0

7.9 Thinking and Attitudes

The Thinking and Attitudes scale is derived from the Criminal Thinking Self-Report scale in the standard COMPAS Core. In the development sample, there were 15,641 individuals with completed Criminal Thinking Self-Report scales. In the analysis, the decision was made to remove the answer choice "Not Sure/Neutral" from all nine retained items. This reduced the number of individuals in the sample to 9,682.

Review of Theoretical and Substantive Validity

Procriminal Attitudes is identified by Andrews and Bonta (2016) as one of the Central Eight risk/need factors. According to those authors:

> 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).

Scale Items

The Thinking and Attitudes scale is constructed using nine items from the Personality/Attitudes domain of the COMPAS-R Core. These items and their short descriptions are shown in Table 91.

The Thinking and Attitudes scale items focus primarily on cognitions including moral justification of criminal behavior, refusal to accept responsibility, victim-blaming, and other rationalizations that minimize the seriousness of crime.

Table 91: COMPAS-R Core Thinking and Attitudes Scale Items. The respondent is instructed that the following statements are about their feelings and beliefs about various things. They are asked to indicate how much they agree or disagree with each of the statements, and that there are no right or wrong answers.

Item	Short Description (Response Categories)
dirt_R	Some people don't deserve any respect and should be treated like dirt. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*
hungry_steal_R	A hungry person has a right to steal. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*
insults_R	If someone insults my friends, family or group they are asking for trouble. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*
law_doesnt_R	The law doesn't help average people. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*
law_nochance_R	When people get into trouble with the law it's because they have no chance to get a decent job. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*
roughly_R	Some people must be treated roughly or beaten up just to send them a clear message. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*
stolen_rich_R	When things are stolen from rich people they won't miss the stuff because insurance will cover the loss. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*
threaten_R	I wouldn't hesitate to hit or threaten people if they have done something to hurt my friends or family. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*
trouble_no_R	Many people get into trouble or use drugs because society has given them no education, jobs or future. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*

Item changes from the Criminal Thinking Self-Report scale in the standard COMPAS Core

Only one item on the Criminal Thinking Self-Report scale from the standard COMPAS Core was removed. It asked the respondent the degree to which they agreed or disagreed with the statement "When people do minor offenses or use drugs, they don't hurt anyone except themselves."

Additionally, the response "Not Sure/Neutral" was removed from the response choices for all nine retained items. Removal of this answer choice resulted in a reduction of cases: there were 5,959 individuals in the composite norm set who responded that they were Not Sure or Neutral about at least one of the nine retained items, reducing the number in the sample from 15,641 to 9,682.

These changes increased the already high reliability of the scale from 0.824 to 0.895, certainly within an acceptable range of reliability (Cortina 1993). Additional statistics for the scale are included in this section.

Factor structure in the norm data

Table 92: Principal component loadings.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
dirt	0.302	-0.026	0.371	-0.060	0.220	-0.256	-0.372
hungry_steal	0.290	-0.050	0.344	0.175	-0.355	-0.232	0.736
insults	0.329	0.572	-0.513	0.304	-0.299	-0.269	-0.185
law_doesnt	0.339	-0.273	-0.283	-0.743	-0.412	0.050	-0.044
law_nochance	0.336	-0.293	0.044	0.458	-0.269	0.683	-0.209
roughly	0.341	0.119	0.309	-0.100	0.198	-0.003	-0.227
stolen_rich	0.334	-0.057	0.290	0.048	-0.115	-0.294	-0.200
threaten	0.367	0.492	0.003	-0.241	0.408	0.449	0.340
trouble_no	0.354	-0.500	-0.469	0.203	0.525	-0.225	0.183

Table 93: Eigenvalues and the proportion of the variance explained by each principal component.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Eigenvalue	1.786	0.316	0.255	0.214	0.191	0.155	0.140
Standard deviation	1.336	0.562	0.505	0.463	0.437	0.393	0.375
Proportion of variance	0.546	0.097	0.078	0.066	0.058	0.047	0.043
Cumulative proportion	0.546	0.642	0.720	0.786	0.844	0.892	0.935

Principal components analysis of the Thinking and Attitudes items in the COMPAS-R composite norm group is presented here. The loadings of the first seven principal components are shown in Table 92. Table 93 shows the eigenvalues of the principal components and the proportion of the variance explained by each. As shown in the table, the first principal component has an eigenvalue of 1.786 and explains 54.6% of the variance in the data. Seven principal components are required to explain st least 90% of the variation in the data.

Scale consistency and reliability in the norm data

Table 94: Correlation of items with scale score, and alpha reliability of the scale when each item is removed. Compare with alpha of 0.895 for all nine items.

	Correlation with score	Alpha when item removed
dirt	0.753	0.882
hungry_steal	0.717	0.885
insults	0.678	0.890
law_doesnt	0.714	0.886
law_nochance	0.742	0.883
roughly	0.809	0.877
stolen_rich	0.789	0.878
threaten	0.752	0.883
trouble_no	0.713	0.887

Cronbach's alpha measures a scale's internal consistency and reliability. Alpha for the Thinking and Attitudes scale in the composite norm set is 0.895. Table 94 shows the correlation of each item with the overall scale score, and the value of alpha when the item is removed from the scale. Removing any of the nine items would lower alpha.

Scale distributions in the norm groups

Table 95: Tabulation by population and gender with mean scores.

Group	N	Mean
Jail Female	214	7.7
Prison Female	662	5.7
Probation Female	1146	6.1
Jail Male	945	9.0
Prison Male	3313	7.2
Probation Male	3402	7.0
ALL	9682	7.1

Table 95 shows the mean Thinking and Attitudes scale score for each population subgroup in the normative data. The means scores for Men in the Prison and Probation groups are comparable, as are those for Women in the Prison and Probation groups. Men and Women in the Jail group are generally higher.

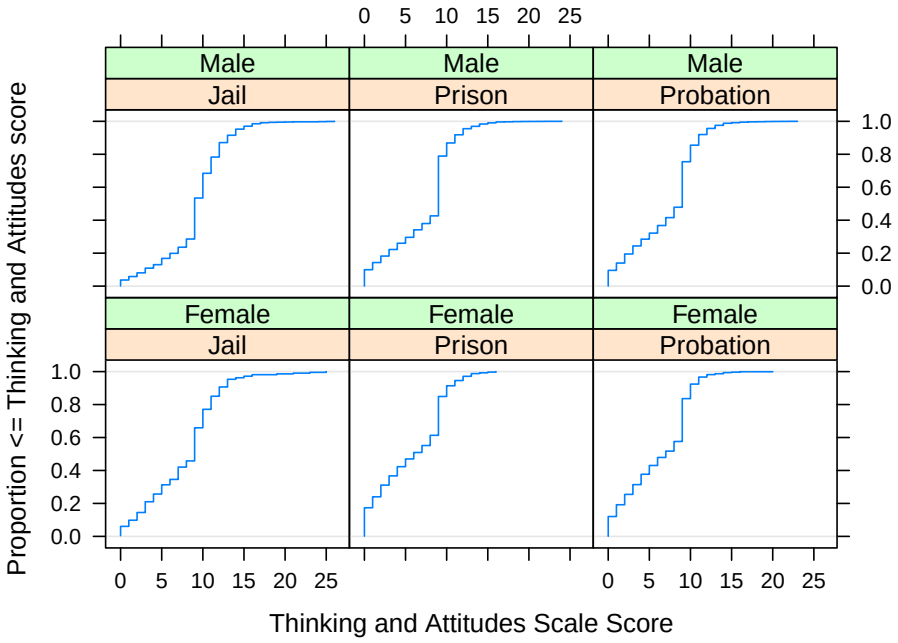


Figure 17: Cumulative distribution plots comparing subgroups in the norm data on the Thinking and Attitudes scale

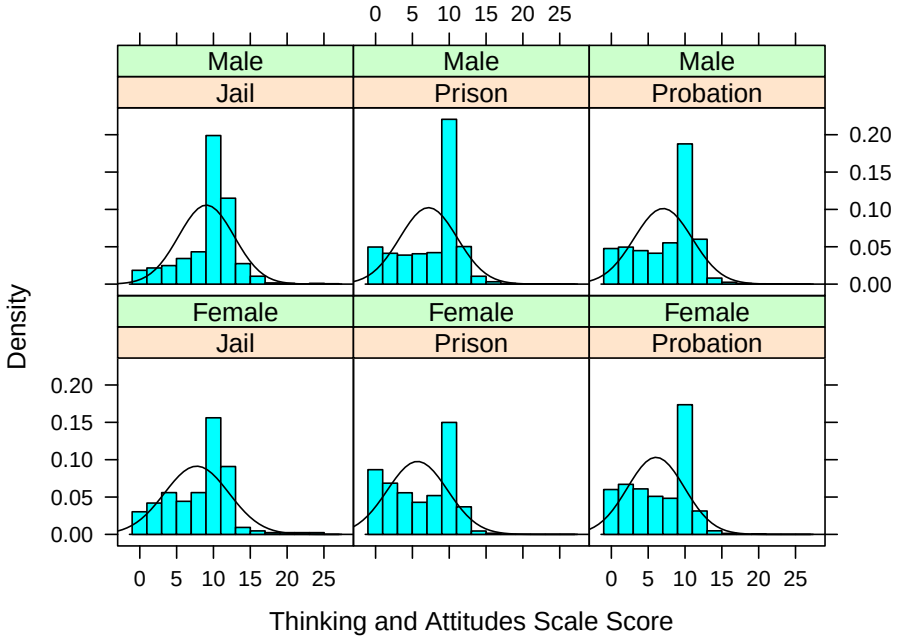


Figure 18: Histograms comparing subgroups in the norm data on the Thinking and Attitudes scale

The cumulative distributions for the subgroups are shown in Figure 17 and the histograms are shown in Figure 18. These plots enhance the information in Table 95, and also show that the score

Table 96: Score ranges for the Thinking and Attitudes scale

	Level		
	Unlikely	Probable	Highly Probable
GN Composite	0 – 9		10 – 27
GN Prison	0 – 9		10 – 27
GN Probation	0 – 8	9	10 – 27
GN Jail (Client-specific)	0 – 9	10	11 – 27
Male Composite	0 – 9		10 – 27
Male Prison	0 – 9		10 – 27
Male Probation	0 – 9		10 – 27
Male Jail (Client-specific)	0 – 9	10 – 11	12 – 27
Female Composite	0 – 7	8 – 9	10 – 27
Female Prison	0 – 6	7 – 9	10 – 27
Female Probation	0 – 7	8 – 9	10 – 27
Female Jail (Client-specific)	0 – 9	10	11 – 27

mode for every group is 9 or 10. Most groups have scores that mainly cluster below the mode, with only the Men in Jail group containing a substantial proportion who scored above 10.

Score ranges

Table 96 displays the score ranges for each of the subgroups in the norm data. GN is a subgroup that includes all the Men and Women. Since there are more Men in the data, the GN ranges more strongly reflect the ranges for Men than for Women. Most agencies choose to use ranges based on gender-specific norms; GN is not available in the software at this time.

The ranges are determined by splitting the scores of each subgroup into the lowest 50% of scores for Unlikely, the highest 30% of scores for Highly Probable, and the remaining 20% in between for Probable. This table of ranges reflects what was seen earlier in the score distributions by subgroup. Since the Men's scores have such a high concentration closer to the top of the score range, there is a sharp break between those with Unlikely scores and those with Highly Probable scores. Women's scores tend to be more evenly distributed, so more of them obtain Probable scores.

Mokken Scale Analysis (MSA): Checking the observable measurement properties of the Monotone Homogeneity Model

Item-pair scalability coefficient The final Thinking and Attitudes scale variables are listed in Table 91. Table 97 displays the item-pair scalability coefficients for all Thinking and Attitudes scale items. All values are positive with the smallest scalability coefficient (0.446) between ***trouble_no and threaten***. The largest scalability coefficient is 0.76 between ***roughly and threaten***.

Table 97: Item-pair scalability coefficients (H_{ij}). Higher H_{ij} result in stronger scales; $-\infty < H_{ij} \leq 1$. The monotone homogeneity (IRT) model assumption requires that $0 \leq H_{ij} \leq 1$.

	hungry_steal	law_doesnt	law_nochance	trouble_no	dirt	roughly	insults	threaten	stolen_rich
hungry_steal		0.519	0.550	0.474	0.660	0.619	0.539	0.493	0.621
law_doesnt	0.519		0.539	0.543	0.640	0.650	0.455	0.503	0.634
law_nochance	0.550	0.539		0.600	0.634	0.624	0.531	0.447	0.620
trouble_no	0.474	0.543	0.600		0.627	0.591	0.457	0.446	0.593
dirt	0.660	0.640	0.634	0.627		0.738	0.575	0.664	0.701
roughly	0.619	0.650	0.624	0.591	0.738		0.679	0.760	0.673
insults	0.539	0.455	0.531	0.457	0.575	0.679		0.700	0.624
threaten	0.493	0.503	0.447	0.446	0.664	0.760	0.700		0.583
stolen_rich	0.621	0.634	0.620	0.593	0.701	0.673	0.624	0.583	

* The standard errors (se) have been removed to allow all item names to fit in the table. The standard errors range from 0.011 to 0.015. The smallest item-pair scalability coefficient is 0.446 and the largest is 0.76.

Item-scalability coefficient The item scalability coefficients are in Table 98; they are all above the lower bound of 0.3 as desired.

Test Scalability Coefficient The test scalability coefficient (H) for the Thinking and Attitudes scale in its final form is 0.585 which is within the limits for the monotone homogeneity model. A scalability coefficient of $H = 0.585$ implies that the scale is weak. The standard error for the statistic is (0.006). The original data had a test scalability coefficient of 0.378 with a standard error of (0.004).

Differential Item Functioning (DIF) Differential Item Functioning (DIF) was tested on a subset of the data ($n = 1000$) after variable transformations for race; the *Other* category for race was not included since it only made up 7% of the final sample. Table 99 denotes the ethnicity proportions from the random subset of $n = 1000$ from the transformed data.

Table 98: Mokken Analysis: Item H values should be above 0.3

	Item H	se
hungry_steal	0.555	(0.008)
law_doesnt	0.552	(0.009)
law_nochance	0.563	(0.008)
trouble_no	0.536	(0.008)
dirt	0.657	(0.008)
roughly	0.667	(0.007)
insults	0.563	(0.009)
threaten	0.565	(0.008)
stolen_rich	0.629	(0.008)

Table 99: Percent of Black, Hispanic, and White persons in a random sample of 1,000 cases drawn from the composite norm data; the data includes all variable transformations (n=9,682).

Ethnicity	Percent
Black	44
Hispanic	14
White	43

The following items were found in the random subsample of $n = 1{,}000$ to have differential item functioning (DIF): ***hungry_steal***. In the original data, a random subsample of $n = 1{,}000$ was also selected; the following variables were flagged with DIF: ***hungry_steal and minor_dont***.

Non-Parametric Monotone Homogeneity (IRT) Model

The three assumptions for the monotone homogeneity model are the *Unidimensionality*, *Local Independence*, and *Latent Monotonicity*.

Unidimensionality The Unidimensionality assumption is tested using Principal Components Analysis (PCA). This was done in section 7.9. We see in Table 93 that the first eigenvalue (1.786) is much larger than the remaining six and over fifty percent (54.6%) of the variance is explained. This is evidence that the *Thinking and Attitudes* scale is unidimensional.

Local Independence The assumption of local independence was tested using the Chi-Squared Test of Independence in R with the parameter `simulate.p.value` set to `TRUE`. All p-values were found to be less than 0.001, indicating evidence of local dependence.

Latent Monotonicity The latent monotonicity assumption is investigated by examining the observable property of the test data known as manifest monotonicity. We see in Table 14 that there are several significant (#zsig) violations (#vi).

Table 100: Manifest monotonicity results: used to assess latent monotonicity. The function `check.mononotinicity` was used from the R package *mokken*.

	ItemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
hungry_steal	0.55	44	0	0.00	0.00	0.00	0.0000	0.00	0	0
law_doesnt	0.55	45	1	0.02	0.05	0.05	0.0011	4.59	1	33
law_nochance	0.56	45	0	0.00	0.00	0.00	0.0000	0.00	0	0
trouble_no	0.54	45	2	0.04	0.05	0.10	0.0022	5.27	2	49
dirt	0.66	45	2	0.04	0.04	0.07	0.0015	3.30	2	30
roughly	0.67	22	0	0.00	0.00	0.00	0.0000	0.00	0	0
insults	0.56	45	1	0.02	0.08	0.08	0.0018	6.26	1	46
threaten	0.56	63	2	0.03	0.05	0.08	0.0013	5.55	2	45
stolen_rich	0.63	30	0	0.00	0.00	0.00	0.0000	0.00	0	0

7.10 Personality Traits Scale and Anger Scale

The Personality Traits scale is derived from the Criminal Personality scale in the standard COMPAS Core. In the development sample, there were 16,097 individuals with completed Criminal Personality scales. In the analysis, the decision was made to remove the answer choice "Not Sure/Neutral" from all nine retained items. This reduced the number of individuals in the sample to 8,766.

Review of Theoretical and Substantive Validity

As stated in the original psychometric studies of COMPAS (Brennan, Dieterich, and Oliver 2007), personality traits related to recidivism include impulsivity, risk-taking, restlessness and boredom, absence of guilt (callousness), selfishness and narcissism, interpersonal dominance, anger and hostility, and a tendency to exploit others (Hare 1991; Cooke, Forth, and Hare 1998), traits this scale seeks to measure.

Scale Items

The Personality Traits scale is constructed using ten items from the Personality/Attitudes domain of the COMPAS-R Core. These items and their short descriptions are shown in Table 101.

The Anger scale is a subset of items from the Personality Traits scale. All five of the items from the Anger scale in the standard COMPAS Core are retained in the COMPAS-R Core Anger scale, and are scored the same way. Its items and their short descriptions are shown in Table 102.

The analyses that follow apply only to the Personality Traits scale, but we can conclude that the Anger scale is working reliably since its items load heavily on the first principal component (see Table 103).

Table 101: COMPAS-R Core Personality Traits Scale Items. In this self-report scale, the respondent is instructed that the next few statements are about what they are like as a person, what their thoughts are, and how other people see them. There are no right or wrong answers. They should just indicate how much they agree or disagree with each statement.

Item	Short Description (Response Categories)
angry_dangerous_R	If people make me lose my temper, I can be dangerous. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*
cold_R	Others see me as cold and unfeeling. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*
getting_close_R	The trouble with getting close to people is that they make demands on you. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*
never_temper_R	I never lose my temper. *(Strongly Disagree=3; Disagree=2; Agree=1; Strongly Agree=0)*
often_bored_R	I am often restless and bored. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*
short_temper_R	I have a short temper and can get angry quickly. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*
sweet_talk_R	I have the ability to sweet talk people to get what I want. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*
talking_problems_R	I am really good at talking my way out of problems. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*
violent_person_R	Some people see me as a violent person. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*
without_thinking_R	I have gotten into trouble because I have acted without thinking. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*

Table 102: COMPAS-R Core Anger Scale Items. In this self-report scale, the respondent is instructed that the next few statements are about what they are like as a person, what their thoughts are, and how other people see them. There are no right or wrong answers. They should just indicate how much they agree or disagree with each statement.

Item	Short Description (Response Categories)
angry_dangerous_R	If people make me lose my temper, I can be dangerous. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*
cold_R	Others see me as cold and unfeeling. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*
never_temper_R	I never lose my temper. *(Strongly Disagree=3; Disagree=2; Agree=1; Strongly Agree=0)*
short_temper_R	I have a short temper and can get angry quickly. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*
violent_person_R	Some people see me as a violent person. *(Strongly Disagree=0; Disagree=1; Agree=2; Strongly Agree=3)*

Item changes from the Criminal Personality scale in the standard COMPAS Core to the Personality Traits scale in COMPAS-R Core

The three items from the standard COMPAS Core Criminal Personality scale that were removed are the following, that ask how much the person agrees or disagrees that

- they feel bad if they break a promise they have made
- they get involved in things they wish they could get out of, and
- to get ahead in life, you must always put yourself first.

Additionally, the response "Not Sure/Neutral" was removed from the response choices for all ten retained items. Removal of this answer choice resulted in a reduction of cases: there were 7,331 individuals in the composite norm set who responded that they were Not Sure or Neutral about at least one of the ten retained items, reducing the number in the sample from 16,097 to 8,766.

These changes increased the reliability of the scale from 0.733 to 0.809, remaining within an acceptable range of reliability (Cortina 1993).

Additional statistics for the Personality Traits scale are included in this section.

Factor structure in the norm data

Table 103: Principal component loadings.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
angry_dangerous	0.370	-0.140	0.032	-0.190	0.238	-0.349	0.324	-0.301
cold	0.320	-0.109	0.068	0.064	0.320	0.185	-0.684	0.359
getting_close	0.294	0.046	0.094	0.204	0.283	0.781	0.364	-0.153
never_temper	0.242	-0.594	-0.384	0.065	-0.604	0.207	-0.082	-0.106
often_bored	0.318	0.293	-0.093	0.827	-0.103	-0.316	-0.027	-0.060
short_temper	0.379	-0.191	-0.080	-0.075	0.074	-0.190	0.446	0.597
sweet_talk	0.270	0.112	0.490	-0.106	-0.318	0.019	-0.084	-0.396
talking_problems	0.238	0.104	0.572	-0.132	-0.383	-0.002	-0.001	0.343
violent_person	0.364	-0.205	-0.028	-0.191	0.315	-0.211	-0.276	-0.334
without_thinking	0.330	0.652	-0.505	-0.399	-0.183	0.106	-0.075	0.020

Table 104: Eigenvalues and the proportion of the variance explained by each principal component.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Eigenvalue	1.834	0.569	0.541	0.436	0.417	0.318	0.232	0.189
Standard deviation	1.354	0.754	0.735	0.660	0.646	0.564	0.482	0.434
Proportion of variance	0.375	0.116	0.111	0.089	0.085	0.065	0.048	0.039
Cumulative proportion	0.375	0.492	0.602	0.691	0.777	0.842	0.889	0.928

Principal components analysis of the Personality Traits items in the COMPAS-R composite norm group is presented here. The loadings of the first eight principal components are shown in Table 103. Table 104 shows the eigenvalues of the first eight principal components and the proportion of the variance explained by each. As shown in the table, the first principal component has an eigenvalue

of 1.834 and explains 37.5% of the variance in the data. Eight principal components are required to explain at least 90% of the variation in the data.

Scale consistency and reliability in the norm data

Table 105: Correlation of items with scale score, and alpha reliability of the scale when each item is removed. Compare with alpha of 0.809 for all ten items.

	Correlation with score	Alpha when item removed
angry_dangerous	0.720	0.777
cold	0.670	0.784
getting_close	0.601	0.793
never_temper	0.464	0.813
often_bored	0.569	0.800
short_temper	0.741	0.774
sweet_talk	0.591	0.794
talking_problems	0.525	0.802
violent_person	0.707	0.779
without_thinking	0.537	0.807

Cronbach's alpha measures a scale's internal consistency and reliability. Alpha for the Personality Traits scale in the composite norm set is 0.809. Table 105 shows the correlation of each item with the overall scale score, and the value of alpha when the item is removed from the scale.

Scale distributions in the norm groups

Table 106: Tabulation by population and gender with mean scores.

Group	N	Mean
Jail Female	205	12.2
Prison Female	583	9.6
Probation Female	1055	9.5
Jail Male	846	12.2
Prison Male	2909	10.1
Probation Male	3168	10.0
ALL	8766	10.2

Table 106 shows the mean Personality Traits scale score for each population subgroup in the normative data. Within corrections populations, cores for Men and Women are about the same on average, with Jail scores higher than for other corrections populations.

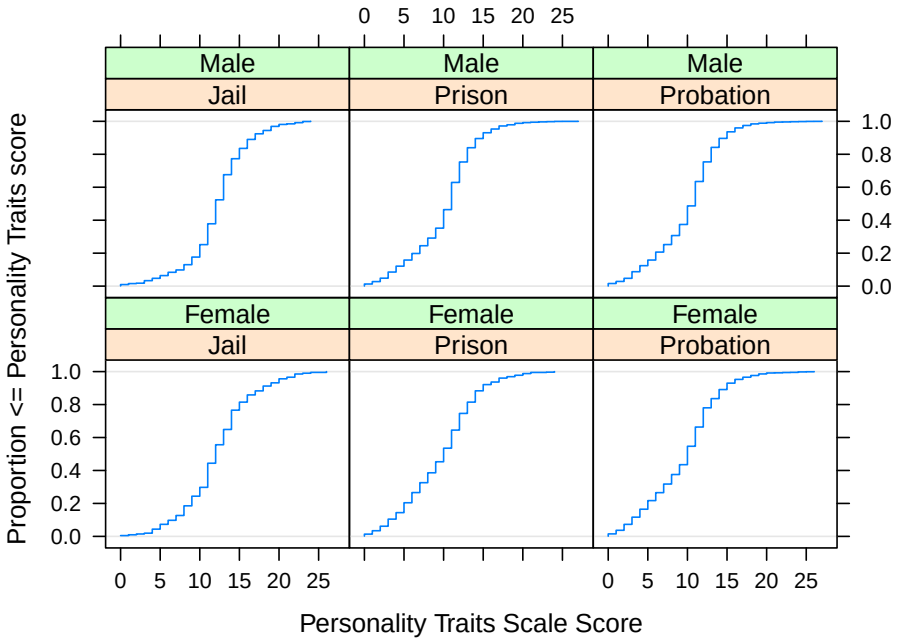


Figure 19: Cumulative distribution plots comparing subgroups in the norm data on the Personality Traits scale

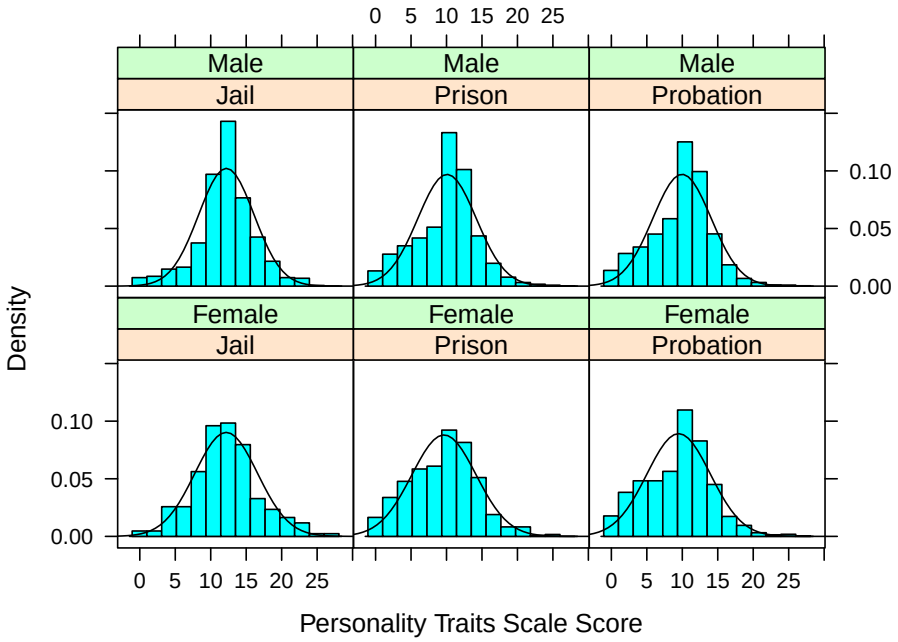


Figure 20: Histograms comparing subgroups in the norm data on the Personality Traits scale

The cumulative distributions for the subgroups are shown in Figure 19, and the histograms for the subgroups are shown in Figure 20. These plots show the scores are concentrated near the means,

Table 107: Score ranges for the Personality Trait scale

	Level		
	Unlikely	Probable	Highly Probable
GN Composite	0 – 11	12	13 – 30
GN Prison	0 – 11	12	13 – 30
GN Probation	0 – 10	11 – 12	13 – 30
GN Jail (Client-specific)	0 – 12	13 – 14	15 – 30
Male Composite	0 – 11	12	13 – 30
Male Prison	0 – 11	12	13 – 30
Male Probation	0 – 11	12	13 – 30
Male Jail (Client-specific)	0 – 12	13 – 14	15 – 30
Female Composite	0 – 10	11 – 12	13 – 30
Female Prison	0 – 10	11 – 12	13 – 30
Female Probation	0 – 10	11 – 12	13 – 30
Female Jail (Client-specific)	0 – 12	13 – 14	15 – 30

with Jail scores somewhat higher on average than in the other populations.

Score ranges for the Personality Traits scale

Tables 107 and 108 show the score ranges for each of the subgroups in the norm data. GN is a subgroup that includes all the Men and Women. Since there are more Men in the data, the GN ranges more strongly reflect the ranges for Men than for Women. Most agencies choose to use ranges based on gender-specific norms; GN is not available in the software at this time.

The ranges are determined by splitting the scores of each subgroup into the lowest 50% of scores for Unlikely, the highest 30% of scores for Highly Probable, and the remaining 20% in between for Probable. Score ranges for Men and Women are very similar.

Score ranges for the Anger scale

The ranges for the Anger scale are determined by splitting the scores of each subgroup into the lowest 50% of scores for Unlikely, the highest 30% of scores for Highly Probable, and the remaining 20% in between for Probable. Score ranges for Men and Women are very similar.

Table 108: Score ranges for the Anger scale

	Level		
	Unlikely	Probable	Highly Probable
GN Composite	0 – 5	6	7 – 15
GN Prison	0 – 5	6	7 – 15
GN Probation	0 – 5	6	7 – 15
GN Jail (Client-specific)	0 – 5		6 – 15
Male Composite	0 – 5	6	7 – 15
Male Prison	0 – 5	6	7 – 15
Male Probation	0 – 5	6	7 – 15
Male Jail (Client-specific)	0 – 5		6 – 15
Female Composite	0 – 4	5	6 – 15
Female Prison	0 – 4	5	6 – 15
Female Probation	0 – 4	5	6 – 15
Female Jail (Client-specific)	0 – 5	6	7 – 15

Mokken Scale Analysis (MSA): Checking the observable measurement properties of the Monotone Homogeneity Model

Item-pair scalability coefficient The final Personality and Attitudes scale variables are listed in Table 101. Table 109 displays the item-pair scalability coefficients for all Personality and Attitudes scale items. All values are positive with the smallest scalability coefficient (0.111) between *talking_problems and never_temper*. The largest scalability coefficient is 0.64 between *angry_dangerous and violent_person*.

Table 109: Item-pair scalability coefficients (H_{ij}). Higher H_{ij} result in stronger scales; $-\infty < H_{ij} \leq 1$. The monotone homogeneity (IRT) model assumption requires that $0 \leq H_{ij} \leq 1$.

	without_thinking	often_bored	cold	getting_close	sweet_talk	talking_problems	angry_dangerous	violent_person	never_temper	short_temper
without_thinking		0.359	0.271	0.305	0.266	0.192	0.363	0.336	0.139	0.392
often_bored	0.359		0.402	0.340	0.278	0.219	0.354	0.324	0.193	0.394
cold	0.271	0.402		0.569	0.405	0.328	0.488	0.553	0.293	0.521
getting_close	0.305	0.340	0.569		0.291	0.241	0.414	0.440	0.196	0.423
sweet_talk	0.266	0.278	0.405	0.291		0.578	0.425	0.379	0.153	0.367
talking_problems	0.192	0.219	0.328	0.241	0.578		0.362	0.314	0.111	0.322
angry_dangerous	0.363	0.354	0.488	0.414	0.425	0.362		0.640	0.332	0.627
violent_person	0.336	0.324	0.553	0.440	0.379	0.314	0.640		0.375	0.621
never_temper	0.139	0.193	0.293	0.196	0.153	0.111	0.332	0.375		0.488
short_temper	0.392	0.394	0.521	0.423	0.367	0.322	0.627	0.621	0.488	

* The standard errors (se) have been removed to allow all item names to fit in the table. The standard errors range from 0.011 to 0.016. The smallest item-pair scalability coefficient is 0.111 and the largest is 0.64.

Item-scalability coefficient The item scalability coefficients are in Table 110; they are not all above the lower bound of 0.3.

Table 110: Mokken Analysis: Item H values should be above 0.3

	Item H	se
without_thinking	0.290	(0.009)
often_bored	0.315	(0.008)
cold	0.424	(0.008)
getting_close	0.350	(0.008)
sweet_talk	0.346	(0.008)
talking_problems	0.293	(0.009)
angry_dangerous	0.448	(0.007)
violent_person	0.445	(0.007)
never_temper	0.246	(0.010)
short_temper	0.462	(0.007)

Test Scalability Coefficient The test scalability coefficient (H) for the Personality Traits scale in its final form is 0.360 which is within the limits for the monotone homogeneity model. A scalability coefficient of $H = 0.360$ implies that the scale is weak. The standard error for the statistic is (0.005). The original data had a test scalability coefficient of 0.211 with a standard error of (0.003).

Differential Item Functioning (DIF) Differential Item Functioning (DIF) was tested on a subset of the data ($n = 1000$) for race, after variable transformations; the *Other* category for race was not included since it only made up 6.9% of the final sample. Table 111 denotes the ethnicity proportions from the random subset of $n = 1000$ from the transformed data. One item was flagged: *angry_dangerous*. In the original data, a random subsample of $n = 1{,}000$ was also selected; the following variables were flagged with DIF: *without_thinking and yourself_first*.

Table 111: Percent of Black, Hispanic, and White persons in a random sample of 1,000 cases drawn from the composite norm data; the data includes all variable transformations (n=8766).

Ethnicity	Percent
Black	45
Hispanic	13
White	42

Non-Parametric Monotone Homogeneity (IRT) Model

The three assumptions for the monotone homogeneity model are the *Unidimensionality*, *Local Independence*, and *Latent Monotonicity*.

Unidimensionality The Unidimensionality assumption is tested using Principal Components Analysis (PCA). This was done in section 7.10. We see in Table 104 that the first eigenvalue (1.834) is much larger than the remaining eigenvalues and over thirty percent (37.5%) of the variance is explained. This is moderate evidence that the Personality Traits scale is unidimensional.

Local Independence The assumption of local independence was tested using the Chi-Squared Test of Independence in R with the parameter `simulate.p.value` set to `TRUE`. All p-values were found to be less than 0.001, indicating evidence of local dependence.

Latent Monotonicity The latent monotonicity assumption is investigated by examining the observable property of the test data known as manifest monotonicity. We see in Table 112 that there are several violations (#vi).

Table 112: Manifest monotonicity results: used to assess latent monotonicity. The function `check.mononotinicity` was used from the R package *mokken*.

	ItemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
without_thinking	0.29	63	7	0.11	0.05	0.28	0.0044	5.25	7	93
often_bored	0.32	63	1	0.02	0.05	0.05	0.0008	2.91	1	36
cold	0.42	63	0	0.00	0.00	0.00	0.0000	0.00	0	0
getting_close	0.35	84	0	0.00	0.00	0.00	0.0000	0.00	0	0
sweet_talk	0.35	84	0	0.00	0.00	0.00	0.0000	0.00	0	0
talking_problems	0.29	84	3	0.04	0.04	0.11	0.0013	4.07	3	55
angry_dangerous	0.45	63	0	0.00	0.00	0.00	0.0000	0.00	0	0
violent_person	0.45	63	0	0.00	0.00	0.00	0.0000	0.00	0	0
never_temper	0.25	63	10	0.16	0.12	0.76	0.0121	13.01	10	167
short_temper	0.46	63	0	0.00	0.00	0.00	0.0000	0.00	0	0

Mokken Scale Analysis (MSA): Checking the observable measurement properties of the Monotone Homogeneity Model

Item-pair scalability coefficient The final Anger scale variables are listed in Table 102. Table 113 displays the item-pair scalability coefficients for all Anger scale items. All values are positive. The

smallest scalability coefficient (0.114) is between ***cold and never_temper***. The largest scalability coefficient is 0.519 between ***angry_dangerous and violent_person***.

Table 113: Item-pair scalability coefficients (H_{ij}). Higher H_{ij} result in stronger scales; $-\infty < H_{ij} \leq 1$. The monotone homogeneity (IRT) model assumption requires that $0 \leq H_{ij} \leq 1$.

	cold	se	angry_dangerous	se	violent_person	se	short_temper	se	never_temper	se
cold			0.356	(0.009)	0.431	(0.009)	0.367	(0.010)	0.114	(0.012)
angry_dangerous	0.356	(0.009)			0.519	(0.008)	0.466	(0.009)	0.142	(0.013)
violent_person	0.431	(0.009)	0.519	(0.008)			0.451	(0.009)	0.210	(0.012)
short_temper	0.367	(0.010)	0.466	(0.009)	0.451	(0.009)			0.371	(0.012)
never_temper	0.114	(0.012)	0.142	(0.013)	0.210	(0.012)	0.371	(0.012)		

* The standard errors range from 0.008 to 0.013. The smallest item-pair scalability coefficient is 0.114 and the largest is 0.519.

Item-scalability coefficient The item scalability coefficients are in Table 114; they are not all above the lower bound of 0.3.

Table 114: Mokken Analysis: Item H values should be above 0.3

	Item H	se
cold	0.323	(0.007)
angry_dangerous	0.381	(0.006)
violent_person	0.406	(0.006)
short_temper	0.416	(0.006)
never_temper	0.210	(0.009)

Test Scalability Coefficient The test scalability coefficient (H) for the Anger scale in its final form is 0.350 which is within the limits for the monotone homogeneity model. A scalability coefficient of $H = 0.350$ implies that the scale is very moderate. The standard error for the statistic is (0.006). The original data had a test scalability coefficient of 0.429 with a standard error of (0.005).

Differential Item Functioning (DIF) Differential Item Functioning (DIF) was tested on a subset of the data ($n = 1{,}000$) after variable transformations for race; the *Other* category for race was not included since it only made up 7.3% of the final sample. Table 115 denotes the ethnicity proportions from the random subset of $n = 1{,}000$ from the transformed data.

No items were found to have differential item functioning (DIF) in the $n = 1{,}000$ subsample. In the original data, a subsample of $n = 1{,}000$ was also selected; no items were flagged for DIF in that subset as well.

Table 115: Percent of Black, Hispanic, and White in random sample of 1,000 cases drawn from the composite norm data; the data include all variable transformations.

Ethnicity	Percent
Black	46
Hispanic	14
White	40

Non-Parametric Monotone Homogeneity (IRT) Model

The three assumptions for the monotone homogeneity model are the *Unidimensionality*, *Local Independence*, and *Latent Monotonicity*.

Unidimensionality The Unidimensionality assumption is tested using Principal Components Analysis (PCA). We see in Table 117 that the first eigenvalue is larger (1.779) than the remaining four and 44.1% of the variance is explained within the first eigenvalue. This is good evidence that the Anger scale is unidimensional.

Table 116: Principal component loadings.

	PC1	PC2	PC3	PC4	PC5
cold	0.370	-0.217	0.697	-0.502	-0.279
angry_dangerous	0.468	-0.256	-0.320	0.388	-0.680
violent_person	0.522	-0.221	0.254	0.496	0.607
short_temper	0.503	0.000	-0.573	-0.583	0.283
never_temper	0.344	0.916	0.137	0.109	-0.110

Table 117: Eigenvalues and the proportion of the variance explained by each principal component.

	PC1	PC2	PC3	PC4
Eigenvalue	1.779	0.958	0.486	0.448
Standard deviation	1.334	0.979	0.697	0.669
Proportion of variance	0.441	0.237	0.120	0.111
Cumulative proportion	0.441	0.678	0.798	0.909

Local Independence The assumption of local independence was tested using the Chi-Squared Test of Independence in R with the parameter `simulate.p.value` set to `TRUE`. All p-values were found to be less than 0.001, indicating evidence of local dependence.

Latent Monotonicity The latent monotonicity assumption is investigated by examining the observable property of the test data known as manifest monotonicity. We see in Table 118 that there are several violations (#vi) for ***never_temper***, all which are significant (#zsig).

Table 118: Manifest monotonicity results: used to assess latent monotonicity. The function `check.mononotinicity` was used from the R package *mokken*.

	ItemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
cold	0.32	30	0	0.0	0.00	0.00	0.0000	0.00	0	0
angry_dangerous	0.38	30	0	0.0	0.00	0.00	0.0000	0.00	0	0
violent_person	0.41	30	0	0.0	0.00	0.00	0.0000	0.00	0	0
short_temper	0.42	30	0	0.0	0.00	0.00	0.0000	0.00	0	0
never_temper	0.21	30	6	0.2	0.11	0.42	0.0138	20.13	6	203

7.11 Socialization History Scale*

Both the standard and conviction versions of the Socialization History scale are derived from the respective standard and conviction versions of the Socialization Failure scale in the standard COMPAS Core.[5] In the development sample, there were 7,764 individuals with completed Socialization Failure scales that were used to derive the standard version of the Socialization History scale. 5,165 individuals had completed a conviction version Criminal involvement scale.

Review of Theoretical and Substantive Validity

Socialization problems during childhood and adolescence, particularly within the family, have been linked to crime and delinquency (Lykken 1995). The Socialization Failure scale in the standard COMPAS Core was constructed to include higher order factors that represent early socialization problems, by building on early delinquency, problem behavior in school, inadequate parental socialization, and early drug use (Brennan, Dieterich, and Oliver 2007). That scale attempts to integrate various factors associated with the early onset of behavioral problems and socialization failure in the family. The Socialization History scale aims to replicate the results of the Socialization Failure scale with somewhat fewer items, ten rather than thirteen.

Scale Items

The standard and conviction versions of the Socialization History scale are constructed using ten items from the Family domain of the COMPAS-R Core. These items and their short descriptions are shown in Tables 119 and 120.

The intent of this scale is to identify socialization breakdown through delinquency, family problems, and its early indicators in school. A high score would represent a person who had early behavior problems in school and serious delinquency problems in the community, and whose parent (or parents) had a criminal history or substance use problems.

The analyses that follow apply only to the standard version of the Socialization History scale.

[5] Scales denoted with an asterisk (*) have both standard and conviction versions.

Table 119: COMPAS-R Core Socialization History Scale Items (Standard Version). The respondent is asked to think of events that occurred during their early life and upbringing, involving family life and school experiences, and to answer these as accurately as possible.

Item	Short Description (Response Categories)
conflict_w_teacher_R	How often did you have conflicts with teachers, including arguments, disruptive behavior, etc.? *(Never=0; Sometimes=1; Often=3)*
conv_parent_R	Were any of the adults who raised you ever arrested, that you know of? *(None were ever arrested=0; Only one was ever arrested=1; More than one was arrested=2)*
expelled_R	Were you ever suspended or expelled from school? *(No=0; Yes=1)*
fights_at_school_R	How often did you get in fights while in school? *(Never=0; Sometimes=1; Often=3)*
high_school_R	Did you complete your high school diploma, GED, or equivalent credential, or are you currently enrolled in a school or program to obtain such a credential? *(No=1; Yes=0)*
juv_placement_R	How many prior commitments to a juvenile institution? *(0=0; 1=1; 2 or more=2)*
n_juv_felony_R	How many prior juvenile felony offense arrests? *(0=0; 1=1; 2=2; 3=3; 4=4; 5 or more=5)*
parent_drug_problem_R	Did a parent (or adult who raised you) have a drug or alcohol problem? *(No=0; Yes=1)*
parent_jailed_R	Was a parent (or adult who raised you) ever sent to jail or prison? *(No=0; Yes=1)*
skipped_classes_R	How often did you skip classes while in school? *(Never=0; Sometimes=1; Often=3)*

Table 120: COMPAS-R Core Socialization History Scale Items (Conviction Version). The respondent is asked to think of events that occurred during their early life and upbringing, involving family life and school experiences, and to answer these as accurately as possible.

Item	Short Description (Response Categories)
conflict_w_teacher_R	How often did you have conflicts with teachers, including arguments, disruptive behavior, etc.? *(Never=0; Sometimes=1; Often=3)*
conv_parent_R	Were any of the adults who raised you ever arrested, that you know of? *(None were ever arrested=0; Only one was ever arrested=1; More than one was arrested=2)*
expelled_R	Were you ever suspended or expelled from school? *(No=0; Yes=1)*
fights_at_school_R	How often did you get in fights while in school? *(Never=0; Sometimes=1; Often=3)*
high_school_R	Did you complete your high school diploma, GED, or equivalent credential, or are you currently enrolled in a school or program to obtain such a credential? *(No=1; Yes=0)*
juv_placement_R	How many prior commitments to a juvenile institution? *(0=0; 1=1; 2 or more=2)*
n_juv_felonyconv_R	How many prior juvenile felony adjudications? *(0=0; 1=1; 2=2; 3=3; 4=4; 5 or more=5)*
parent_drug_problem_R	Did a parent (or adult who raised you) have a drug or alcohol problem? *(No=0; Yes=1)*
parent_jailed_R	Was a parent (or adult who raised you) ever sent to jail or prison? *(No=0; Yes=1)*
skipped_classes_R	How often did you skip classes while in school? *(Never=0; Sometimes=1; Often=3)*

Item changes from the Socialization Failure scale in the standard COMPAS Core to the Socialization History scale in the COMPAS-R Core

Two items are not used from the standard COMPAS Core: 1) the item that asks whether someone failed or repeated a grade and 2) the item that asks for the number of juvenile violent felonies. Additionally, the items asking whether a person's father and mother had ever been arrested are joined into a single item asking about the number of arrest events of the person's parents, with answer choices of none, only one, or more than one parent or parent figure (called conv_parent).

These changes decreased the reliability of the scale slightly, from 0.702 to 0.675. Additional statistics for the scale are included in this section.

Factor structure in the norm data

Table 121: Principal component loadings.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
conflict_w_teacher	-0.409	-0.041	0.342	-0.454	0.703	-0.062	0.101
expelled	-0.186	0.023	0.094	-0.059	-0.036	0.252	-0.939
fights_at_school	-0.361	0.079	0.315	-0.492	-0.704	-0.070	0.145
high_school	-0.051	0.010	0.031	-0.005	0.000	0.916	0.254
juv_placement	-0.126	0.299	-0.087	0.019	0.025	0.030	-0.045
n_juv_felony	-0.275	0.835	-0.325	0.057	0.066	-0.044	0.051
parent_drug_problem	-0.102	-0.002	0.207	0.228	-0.018	-0.287	-0.074
parent_jailed	-0.126	0.056	0.308	0.318	-0.010	0.024	0.053
skipped_classes	-0.711	-0.438	-0.482	0.247	-0.057	-0.023	0.064
conv_parent	-0.205	0.104	0.540	0.575	-0.022	0.038	0.078

Table 122: Eigenvalues and the proportion of the variance explained by each principal component.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Eigenvalue	1.675	0.855	0.677	0.575	0.391	0.203	0.177
Standard deviation	1.294	0.925	0.823	0.758	0.625	0.450	0.420
Proportion of variance	0.339	0.173	0.137	0.116	0.079	0.041	0.036
Cumulative proportion	0.339	0.512	0.648	0.765	0.844	0.885	0.921

Principal components analysis of the Socialization History items in the COMPAS-R composite norm group is presented here. The loadings of the first seven principal components are shown in Table 121. Table 122 shows the eigenvalues of the principal components and the proportion of the variance explained by each. As shown in the table, the first principal component has an eigenvalue of 1.675 and explains 33.9% of the variation in the data. The first seven principal components explain 92.1% of the variation in the data.

Scale consistency and reliability in the norm data

Table 123: Correlation of items with scale score, and alpha reliability of the scale when each item is removed. Compare with alpha of 0.675 for all ten items.

	Correlation with score	Alpha when item removed
conflict_w_teacher	0.613	0.630
expelled	0.547	0.642
fights_at_school	0.595	0.634
high_school	0.256	0.680
juv_placement	0.460	0.654
n_juv_felony	0.502	0.669
parent_drug_problem	0.407	0.662
parent_jailed	0.525	0.645
skipped_classes	0.625	0.654
conv_parent	0.554	0.641

Cronbach's alpha measures a scale's internal consistency and reliability. Alpha for the Socialization History scale in the composite norm set is 0.675. Table 123 shows the correlation of each item with the overall scale score, and the value of alpha when the item is removed from the scale. Removing the high school item increases alpha slightly, but it is an important item to have information about. While it is less strongly correlated with the overall scale score than the other items, the correlation is positive and significant. Removing any of the other items would lower alpha.

Scale distributions in the norm groups

Table 124: Tabulation by population and gender with mean scores. Note: this table does not present the scale score from the single transgender person in the data.

Group	N	Mean
Jail Female	404	4.7
Prison Female	394	4.7
Probation Female	756	4.2
Jail Male	1900	4.8
Prison Male	1686	5.2
Probation Male	2623	4.5
ALL	7763	4.8

Table 124 shows the mean Socialization History scale score for each population subgroup in the normative data. Women's and Men's scores are very similar.

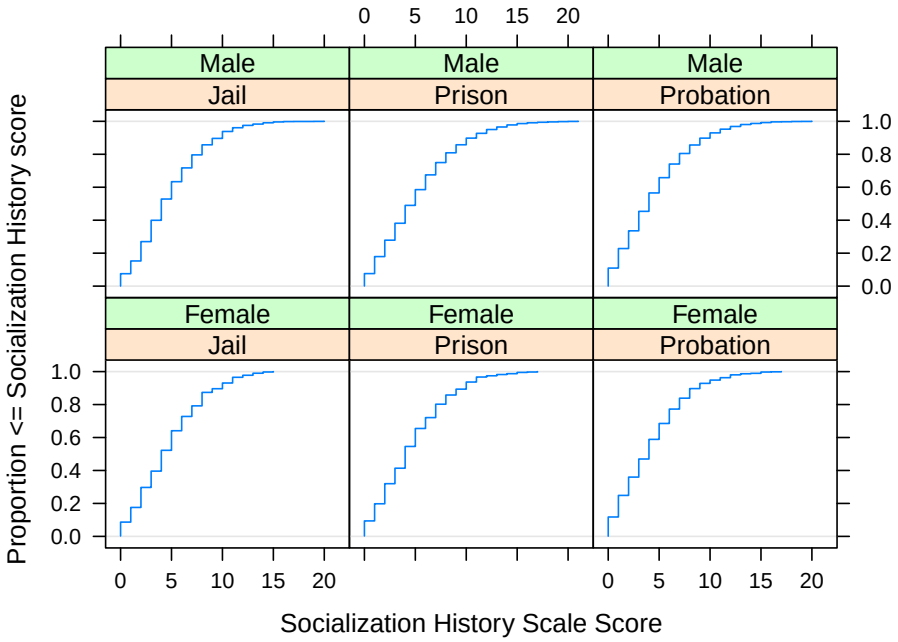


Figure 21: Cumulative distribution plots comparing subgroups in the norm data on the Socialization History scale

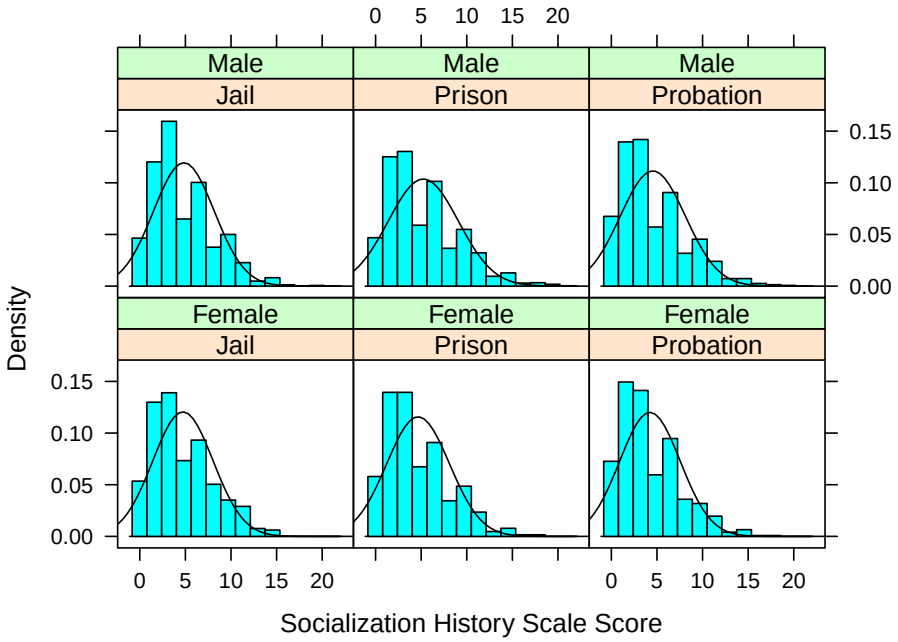


Figure 22: Histograms comparing subgroups in the norm data on the Socialization History scale

The cumulative distributions for the subgroups are shown in Figure 21. These plots enhance the information in Table 124. The distribution of scores for these groups are similar to each other, with

Table 125: Score ranges for the Socialization History scale (standard version)

	Level		
	Unlikely	Probable	Highly Probable
GN Composite	0 – 4	5 – 6	7 – 22
GN Prison	0 – 4	5 – 7	8 – 22
GN Probation	0 – 4	5 – 6	7 – 22
GN Jail (Client-specific)	0 – 4	5 – 6	7 – 22
Male Composite	0 – 4	5 – 6	7 – 22
Male Prison	0 – 5	6 – 7	8 – 22
Male Probation	0 – 4	5 – 6	7 – 22
Male Jail (Client-specific)	0 – 4	5 – 6	7 – 22
Female Composite	0 – 4	5 – 6	7 – 22
Female Prison	0 – 4	5 – 6	7 – 22
Female Probation	0 – 4	5 – 6	7 – 22
Female Jail (Client-specific)	0 – 4	5 – 6	7 – 22

Men scoring somewhat higher than Women; within gender, the distributions across populations are very similar.

The histograms for the subgroups are shown in Figure 22. These plots demonstrate the tendency for somewhat lower scores among Women.

Score ranges

Tables 125 and 126 show the score ranges for each of the subgroups in the norm data. GN is a subgroup that includes all the Men and Women. Since there are more Men in the data, the GN ranges more strongly reflect the ranges for Men than for Women. Most agencies choose to use ranges based on gender-specific norms; GN is not available in the software at this time.

The ranges for both the standard and conviction version of the Socialization History scale are determined by splitting the scores of each subgroup into the lowest 50% of scores for Unlikely, the highest 30% of scores for Highly Probable, and the remaining 20% in between for Probable.

Table 126: Score ranges for the Socialization History scale (conviction version)

	Level		
	Unlikely	Probable	Highly Probable
GN Composite	0 – 3	4 – 5	6 – 22
GN Prison	0 – 4	5	6 – 22
GN Probation	0 – 3	4 – 5	6 – 22
Male Composite	0 – 3	4 – 5	6 – 22
Male Prison	0 – 4	5 – 6	7 – 22
Male Probation	0 – 3	4 – 5	6 – 22
Female Composite	0 – 3	4 – 5	6 – 22
Female Prison	0 – 3	4	5 – 22
Female Probation	0 – 3	4 – 5	6 – 22

Mokken Scale Analysis (MSA): Checking the observable measurement properties of the Monotone Homogeneity Model

Item-pair scalability coefficient The final Socialization History scale variables are listed in Table 119. Table 127 displays the item-pair scalability coefficients for all Socialization History scale items. All values are positive with the smallest scalability coefficient (0.004) between *high_school and parent_drug_problem*. The largest scalability coefficient is 0.881 between *conv_parent and parent_jailed*.

Table 127: Item-pair scalability coefficients (H_{ij}). Higher H_{ij} result in stronger scales; $-\infty < H_{ij} \leq 1$. The monotone homogeneity (IRT) model assumption requires that $0 \leq H_{ij} \leq 1$.

	n_juv_felony	juv_placement	conflict_w_teacher	skipped_classes	fights_at_school	conv_parent	high_school	expelled	parent_jailed	parent_drug_problem
n_juv_felony		0.610	0.147	0.195	0.211	0.154	0.083	0.549	0.175	0.055
juv_placement	0.610		0.173	0.201	0.210	0.165	0.089	0.556	0.203	0.086
conflict_w_teacher	0.147	0.173		0.463	0.409	0.214	0.129	0.673	0.237	0.175
skipped_classes	0.195	0.201	0.463		0.381	0.177	0.105	0.425	0.184	0.205
fights_at_school	0.211	0.210	0.409	0.381		0.188	0.124	0.657	0.213	0.158
conv_parent	0.154	0.165	0.214	0.177	0.188		0.096	0.331	0.881	0.490
high_school	0.083	0.089	0.129	0.105	0.124	0.096		0.204	0.091	0.004
expelled	0.549	0.556	0.673	0.425	0.657	0.331	0.204		0.347	0.232
parent_jailed	0.175	0.203	0.237	0.184	0.213	0.881	0.091	0.347		0.418
parent_drug_problem	0.055	0.086	0.175	0.205	0.158	0.490	0.004	0.232	0.418	

* The standard errors (se) have been removed to allow all item names to fit in the table. The standard errors range from 0.007 to 0.031. The smallest item-pair scalability coefficient is 0.004 and the largest is 0.881.

Table 129: Percent of Black, Hispanic, and White persons in a random sample of 1,000 cases drawn from the composite norm data; the data includes all variable transformations.

Ethnicity	Percent
Black	39
Hispanic	16
White	45

Item-scalability coefficient The item scalability coefficients are in Table 128; most are not above the lower bound of 0.3.

Table 128: Mokken Analysis: Item H values should be above 0.3

	Item H	se
n_juv_felony	0.227	(0.010)
juv_placement	0.272	(0.010)
conflict_w_teacher	0.293	(0.007)
skipped_classes	0.270	(0.008)
fights_at_school	0.286	(0.008)
conv_parent	0.267	(0.007)
high_school	0.101	(0.009)
expelled	0.452	(0.009)
parent_jailed	0.308	(0.008)
parent_drug_problem	0.215	(0.008)

Test Scalability Coefficient The test scalability coefficient (H) for the Socialization History scale in its final form is 0.266 which is within the limits for the monotone homogeneity model. A scalability coefficient of $H = 0.266$ implies that the scale is very weak. The standard error for the statistic is (0.005). The original data had a test scalability coefficient of 0.250 with a standard error of (0.005).

Differential Item Functioning (DIF) Differential Item Functioning (DIF) was tested on a subset of the data ($n = 1{,}000$) for race, after variable transformations; the *Other* category for race was not included since it only made up 8.9% of the final sample. Table 129 denotes the ethnicity proportions from the random subset of $n = 1{,}000$ from the transformed data.

The following items were found to have differential item functioning (DIF) in the $n = 1{,}000$ subsample: ***n_juv_felony, juv_placement, conflict_w_teacher, skipped_classes and parent_jailed***. In the original data, a subsample of $n = 1{,}000$ was also selected; the following variables were flagged with DIF: ***expelled, conflict_w_teacher, skipped_classes and parent_drug_problem***.

Non-Parametric Monotone Homogeneity (IRT) Model

The three assumptions for the monotone homogeneity model are the *Unidimensionality*, *Local Independence*, and *Latent Monotonicity*.

Unidimensionality The Unidimensionality assumption is tested using Principal Components Analysis (PCA). This was done in section 7.11. We see in Table 122 that the first eigenvalue (1.675) is larger than the remaining six and almost twice as large as the second eigenvalue (0.855). However, only 33.9% of the variance is explained within the first eigenvalue. This is moderate evidence that the Socialization History scale is unidimensional.

Local Independence The assumption of local independence was tested using the Chi-Squared Test of Independence in R with the parameter `simulate.p.value` set to `TRUE`. The $n = 1{,}000$ subset used for testing DIF from the transformed data (items changed for the COMPAS-R Core scales) was used. There were 4 pairs of items with p-values greater than or equal to 0.05. These item pairs are listed in Table 130.

Table 130: Item pairs and their corresponding p-value for items that suggest independence from a subset of n = 1,000 from the final COMPAS-R CORE scale. The rows highlighted in black contain the item pairs that were also found to demonstrate independence in the full transformed data. Their respctive p-values were 0.053 and 0.773 for the full sample.

item1	item2	p_value
n_juv_felony	high_school	0.083
n_juv_felony	**parent_drug_problem**	**0.978**
juv_placement	parent_drug_problem	0.850
high_school	**parent_drug_problem**	**0.065**

Latent Monotonicity The latent monotonicity assumption is investigated by examining the observable property of the test data known as manifest monotonicity. We see in Table 131 that there are no violations (#vi).

Table 131: Manifest monotonicity results: used to assess latent monotonicity. The function `check.mononotinicity` was used from the R package *mokken*.

	ItemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
n_juv_felony	0.23	105	0	0	0	0	0	0	0	0
juv_placement	0.27	42	0	0	0	0	0	0	0	0
conflict_w_teacher	0.29	63	0	0	0	0	0	0	0	0
skipped_classes	0.27	63	0	0	0	0	0	0	0	0
fights_at_school	0.29	63	0	0	0	0	0	0	0	0
conv_parent	0.27	42	0	0	0	0	0	0	0	0
high_school	0.10	28	0	0	0	0	0	0	0	0
expelled	0.45	28	0	0	0	0	0	0	0	0
parent_jailed	0.31	21	0	0	0	0	0	0	0	0
parent_drug_problem	0.22	28	0	0	0	0	0	0	0	0

7.12 Financial

The Financial scale is derived from the Financial Problems scale in the standard COMPAS Core. In the development sample, there were 11,204 individuals with completed Financial Problems scales.

Review of Theoretical and Substantive Validity

Difficulty with managing finances is a non-criminogenic need and an example of a stabilization factor that, when addressed prior to treatment for identified criminogenic needs, enhances responsivity to that treatment (Marlowe 2018). The Financial scale focuses on these difficulties, such as the struggle to survive financially, problems paying bills, and other problems related to a shortage of money.

Scale Items

The Financial scale is constructed using four items from the Social Exclusion domain of the COMPAS-R Core. These items and their short descriptions are shown in Table 132.

The Financial scale assesses the degree to which a person experiences and responds to poverty and financial problems, by determining whether the person worries about financial survival, has trouble paying bills, or has conflicts with friends or family over money.

Table 132: COMPAS-R Core Financial Scale Items. In the interview, respondents are asked to think of past work experiences, job experiences, and their financial situation, and to answer as accurately as possible.

Item	Short Description (Response Categories)
fam_conflicts_R	During the last 12 months, how often have you had conflicts with friends or family over money? *(Often=2; Sometimes=1; Never=0)*
get_by_R	During the last 12 months, how often have you had barely enough money to get by? *(Often=2; Sometimes=1; Never=0)*
pay_bills_R	During the last 12 months, how often have you had trouble paying bills? *(Often=2; Sometimes=1; Never=0)*
worry_survive_R	During the last 12 months, how often have you worried about financial survival? *(Often=2; Sometimes=1; Never=0)*

Item changes from the Financial Problems scale in the standard COMPAS Core

Only one item from the standard COMPAS Core was removed to form the Financial Problems scale, the one that asks how often the person worked during the last 12 months at a job that didn't pay over minimum wage. This change increased the reliability of the scale from 0.713 to 0.731, and it remains within an acceptable range of reliability (Cortina 1993). Additional statistics for the scale are included in this section.

Factor structure in the norm data

Table 133: Principal component loadings.

	PC1	PC2	PC3	PC4
fam_conflicts	0.209	-0.780	0.282	0.518
get_by	0.554	0.566	0.492	0.361
pay_bills	0.517	-0.266	0.284	-0.763
worry_survive	0.618	-0.021	-0.773	0.139

Table 134: Eigenvalues and the proportion of the variance explained by each principal component.

	PC1	PC2	PC3	PC4
Eigenvalue	1.146	0.273	0.264	0.230
Standard deviation	1.070	0.523	0.514	0.479
Proportion of variance	0.599	0.143	0.138	0.120
Cumulative proportion	0.599	0.742	0.880	1.000

Principal components analysis of the Financial items in the COMPAS-R composite norm group is presented here. The loadings of the principal components are shown in Table 133. Table 134 shows the eigenvalues of the principal components and the proportion of the variance explained by each. As shown in the table, the first principal component has an eigenvalue of 1.146 and explains 59.9% of the variance in the data.

Scale consistency and reliability in the norm data

Table 135: Correlation of items with scale score, and alpha reliability of the scale when each item is removed. Compare with alpha of 0.731 for all four items.

	Correlation with score	Alpha when item removed
fam_conflicts	0.545	0.768
get_by	0.789	0.636
pay_bills	0.797	0.619
worry_survive	0.814	0.621

Cronbach's alpha measures a scale's internal consistency and reliability. Alpha for the Financial scale in the composite norm set is 0.731. Table 135 shows the correlation of each item with the overall scale score, and the value of alpha when the item is removed from the scale. Removing the family conflicts item would increase alpha, but it is not recommended that a scale have fewer than four items. The family conflicts item is well-correlated with the scale score and alpha is good. Removing any of the other items would lower alpha substantially.

Scale distributions in the norm groups

Table 136: Tabulation by population and gender with mean scores. Note: this table does not present the scale score from the single transgender person in the data.

Group	N	Mean
Jail Female	262	3.6
Prison Female	477	3.1
Probation Female	1306	3.1
Jail Male	1900	3.0
Prison Male	2970	2.4
Probation Male	4288	2.9
ALL	11203	2.9

Table 136 shows the mean Financial scale score for each population subgroup in the normative data. The mean scores for Women are higher than for Men, regardless of which corrections settings are compared .

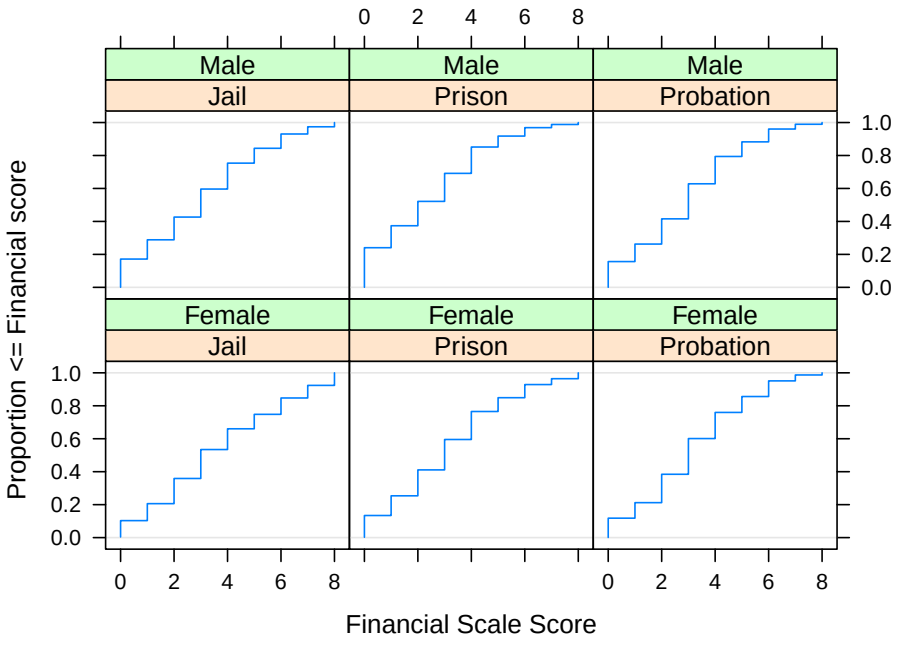


Figure 23: Cumulative distribution plots comparing subgroups in the norm data on the Financial scale

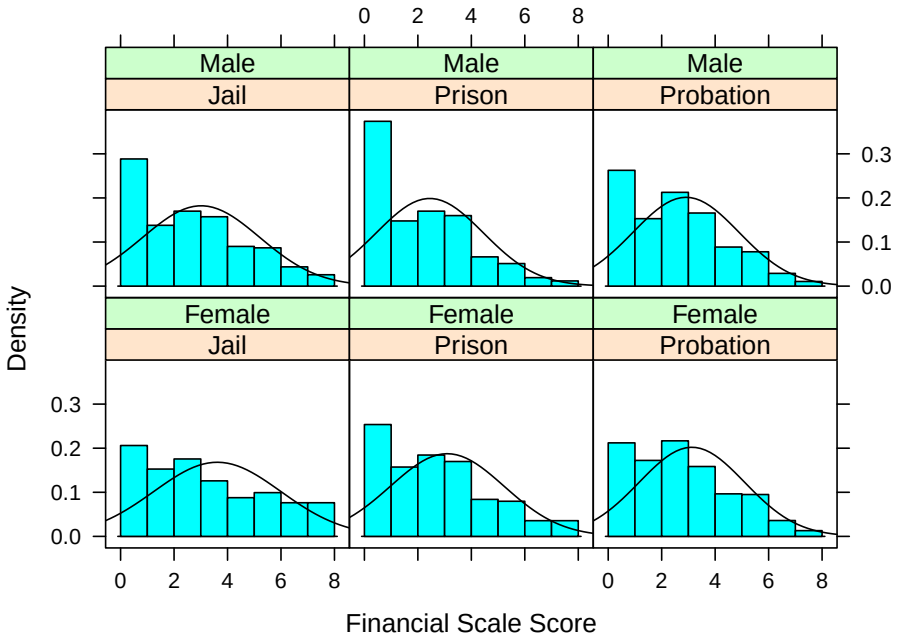


Figure 24: Histograms comparing subgroups in the norm data on the Financial scale

The cumulative distributions for the subgroups are shown in Figure 23 and the histograms for the subgroups are shown in Figure 24. These plots enhance the information in Table 136. For example, the plot for Men Prisoners shows that a higher proportion of them score 0 on this scale than all the other subgroups. Women in Jail, on the other hand, have the largest proportion of higher scores

Score ranges

Table 137 shows the score ranges for each of the subgroups in the norm data. GN is a subgroup that includes all the Men and Women. Since there are more Men in the data, the GN ranges more strongly reflect the ranges for Men than for Women. Most agencies choose to use ranges based on gender-specific norms; GN is not available in the software at this time.

The ranges are determined by splitting the scores of each subgroup into the lowest 50% of scores for Unlikely, the highest 30% of scores for Highly Probable, and the remaining 20% in between for Probable.

Table 137: Score ranges for the Financial scale.

	Level		
	Unlikely	Probable	Highly Probable
GN Composite	0 – 3	4	5 – 8
GN Prison	0 – 2	3 – 4	5 – 8
GN Probation	0 – 3	4	5 – 8
GN Jail (Client-specific)	0 – 3	4	5 – 8
Male Composite	0 – 3	4	5 – 8
Male Prison	0 – 2	3 – 4	5 – 8
Male Probation	0 – 3	4	5 – 8
Male Jail (Client-specific)	0 – 3	4	5 – 8
Female Composite	0 – 3	4	5 – 8
Female Prison	0 – 3	4	5 – 8
Female Probation	0 – 3	4	5 – 8
Female Jail (Client-specific)	0 – 3	4 – 5	6 – 8

Mokken Scale Analysis (MSA): Checking the observable measurement properties of the Monotone Homogeneity Model

Item-pair scalability coefficient The final Financial scale variables are listed in Table 132. Table 138 displays the item-pair scalability coefficients for all Financial scale items. All values are positive with the smallest scalability coefficient (0.303) between ***fam_conflicts and get_by***. The largest scalability coefficient is 0.638 between ***pay_bills and get_by***.

Table 138: Item-pair scalability coefficients (H_{ij}). Higher H_{ij} result in stronger scales; $-\infty < H_{ij} \leq 1$. The monotone homogeneity (IRT) model assumption requires that $0 \leq H_{ij} \leq 1$.

	fam_conflicts	se	worry_survive	se	pay_bills	se	get_by	se
fam_conflicts			0.339	(0.012)	0.409	(0.013)	0.303	(0.012)
worry_survive	0.339	(0.012)			0.624	(0.009)	0.565	(0.009)
pay_bills	0.409	(0.013)	0.624	(0.009)			0.638	(0.009)
get_by	0.303	(0.012)	0.565	(0.009)	0.638	(0.009)		

* The smallest item-pair scalability coefficient is 0.303 and the largest is 0.638.

Item-scalability coefficient The item scalability coefficients are in Table 139; they are all above the lower bound of 0.3.

Table 139: Mokken Analysis: Item H values should be above 0.3

	Item H	se
fam_conflicts	0.348	(0.010)
worry_survive	0.529	(0.006)
pay_bills	0.578	(0.007)
get_by	0.525	(0.007)

Test Scalability Coefficient The test scalability coefficient (H) for the Financial scale in its final form is 0.504 which is within the limits for the monotone homogeneity model. A scalability coefficient of $H = 0.504$ implies that the scale is strong. The standard error for the statistic is (0.006). The original data had a test scalability coefficient of 0.418 with a standard error of (0.006).

Differential Item Functioning (DIF) Differential Item Functioning (DIF) was tested on a subset of the data ($n = 1{,}000$) for race, after variable transformations; the *Other* category for race was not included since it only made up 8.6% of the final sample. Table 140 denotes the ethnicity proportions from the random subset of $n = 1{,}000$ from the transformed data.

The following items were found to have differential item functioning (DIF) in the $n = 1{,}000$ subsample: ***fam_conflicts, worry_survive and pay_bills***. In the original data, a subsample of $n = 1{,}000$ was also selected; the following variables were flagged with DIF: ***welfare***.

Non-Parametric Monotone Homogeneity (IRT) Model

The three assumptions for the monotone homogeneity model are the *Unidimensionality*, *Local Independence*, and *Latent Monotonicity*.

Table 140: Percent of Black, Hispanic, and White persons in a random sample of 1,000 cases drawn from the composite norm data; the data includes all variable transformations.

Ethnicity	Percent
Black	41
Hispanic	17
White	42

Unidimensionality The Unidimensionality assumption is tested using Principal Components Analysis (PCA). This was done in section 7.12. We see in Table 134 that the first eigenvalue (1.146) is much larger than the remaining four and over half (59.9%) of the variance is explained. This is strong evidence that the Financial scale is unidimensional.

Local Independence The assumption of local independence was tested using the Chi-Squared Test of Independence in R with the parameter `simulate.p.value` set to `TRUE`. All p-values were found to be less than 0.001, indicating evidence of local dependence.

Latent Monotonicity The latent monotonicity assumption is investigated by examining the observable property of the test data known as manifest monotonicity. We see in Table 141 that there are no violations (#vi).

Table 141: Manifest monotonicity results: used to assess latent monotonicity. The function `check.mononotinicity` was used from the R package *mokken*.

	ItemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
fam_conflicts	0.35	30	0	0	0	0	0	0	0	0
worry_survive	0.53	20	0	0	0	0	0	0	0	0
pay_bills	0.58	20	0	0	0	0	0	0	0	0
get_by	0.53	20	0	0	0	0	0	0	0	0

7.13 Vocational/Educational

The Vocational/Educational scale is derived from the Vocational/Education scale in the standard COMPAS Core. In the development sample, there were 21,114 individuals with completed Vocational/Education scales.

Review of Theoretical and Substantive Validity

As stated in the 2007 psychometric study of the COMPAS Core scales:

> Another of the "big five" risk factors for crime and recidivism prediction in the Gendreau, Goggin, and Little (1996) meta-analysis is labeled "social achievement." This present concept is an amalgam of educational attainment, vocational skills, job opportunities, a record of stable employment, good income, and, more generally, the level of legitimate economic opportunity. Basically, persons with more social capital have higher "life chances" than other offenders who may have very restricted success opportunities (Hagan 1998; Coleman 1990). The family is of critical importance in the building social capital. Parents either transmit positive and substantial social capital to their child or fail in the socialization processes. This scale is a higher order factor in COMPAS, using items from both educational and vocational domains. Offenders differ greatly in access to social capital or other resources. Social Capital is somewhat dynamic. It can be built or destroyed. For example, a record of serious criminal behavior or high school dropout will clearly diminish life chances and social resources, whereas completing a job skills training course or obtaining a GED may increase these chances.

Andrews and Bonta (2016) include School/Work among the Central Eight recidivism risk factors.

Scale Items

The Vocational/Educational scale is constructed using eight items from the Social Exclusion domain of the COMPAS-R Core. These items and their short descriptions are shown in Table 142.

The Vocational/Educational scale items focus primarily on a person's employment status both current and recent. Several of the items implicitly include past educational achievement, such as whether one has a skill, trade, or profession at which they usually find work or whether they need more job training. The item from the standard COMPAS Core that asks whether the person has a high school diploma has been modified to ask if the person completed their high school diploma, GED, or equivalent credential, *or* if they are currently enrolled in a school or program to obtain such a credential. This modification allows a younger person who is in the process of completing their secondary education to avoid incurring an inflated Vocational/Educational need score.

The Vocational/Educational scale score is also an input for the Summative General Recidivism Risk Scale.

Table 142: COMPAS-R Core Vocational/Educational Scale Items. The interviewer is instructed to ask the respondent to think of their past school and work experiences.

Item	Short Description (Response Categories)
chance_success_work_R	Right now, if you were to get (or have) a good job how would you rate your chance of being successful? *(Good=0; Fair=1; Poor=2)*
haveempschool_R	Do you have a verifiable employer or school? *(No=1; Yes=0)*
high_school_R	Did you complete your high school diploma, GED, or equivalent credential, or are you currently enrolled in a school or program to obtain such a credential? *(No=1; Yes=0)*
job_last_year_R	How much have you worked or been enrolled in school within the last 12 months? *(12 months full time=0; 12 months part time=1; More than 6 months (up to, but not including a year), either part time or full time=2; Between 0 and 6 months, either part time or full time=3)*
job_R	Do you currently have a job? *(No=1; Yes=0)*
need_training_R	Right now, do you feel you need more training in a new job or career skill? *(No=0; Yes=1)*
skill_R	Do you have a skill, trade or profession in which you usually find work? *(No=1; Yes=0)*
wages_above_min_R	How hard is it for you to find a job above minimum wage compared to others? *(Easier=0; Same=1; Harder=2; Much harder=3)*

Item changes from the Vocational/Education scale in the standard COMPAS Core

It was found that the following items from the Vocational/Education scale in the standard COMPAS Core could be removed without impacting reliability:

1. the item asking about usual grades in high school,
2. the item asking whether the person had ever failed or repeated a grade, and
3. the item asking whether the person had ever been expelled from school.

These changes had no impact on the alpha reliability of the scale, with $\alpha = 0.671$.

Additional statistics for the scale are included in this section.

Factor structure in the norm data

Table 143: Principal component loadings.

	PC1	PC2	PC3	PC4	PC5
chance_success_work	0.096	-0.130	0.356	0.236	0.643
haveempschool	0.209	0.006	0.544	-0.007	-0.387
high_school	0.053	-0.061	0.254	-0.017	0.459
job	0.222	0.010	0.513	-0.009	-0.387
job_last_year	0.863	0.409	-0.274	0.018	0.109
need_training	0.064	-0.121	-0.011	-0.953	0.117
skill	0.121	-0.101	0.352	-0.142	0.191
wages_above_min	0.362	-0.887	-0.225	0.121	-0.120

Table 144: Eigenvalues and the proportion of the variance explained by each principal component.

	PC1	PC2	PC3	PC4	PC5
Eigenvalue	2.131	0.818	0.298	0.237	0.223
Standard deviation	1.460	0.905	0.546	0.486	0.472
Proportion of variance	0.518	0.199	0.072	0.058	0.054
Cumulative proportion	0.518	0.717	0.789	0.846	0.901

Principal components analysis of the Vocational/Educational items in the COMPAS-R composite norm group is presented here. The loadings of the first five principal components are shown in Table 143. Table 144 shows the eigenvalues of the first five principal components and the proportion of the variance explained by each. As shown in the table, the first principal component has an eigenvalue of 2.131 and explains 51.8% of the variance in the data. Five principal components are required to explain at least 90% of the variation in the data.

Scale consistency and reliability in the norm data

Table 145: Correlation of items with scale score, and alpha reliability of the scale when each item is removed. Compare with alpha of 0.671 for all eight items.

	Correlation with score	Alpha when item removed
chance_success_work	0.441	0.656
haveempschool	0.689	0.607
high_school	0.328	0.674
job	0.708	0.602
job_last_year	0.780	0.637
need_training	0.340	0.674
skill	0.536	0.639
wages_above_min	0.662	0.627

Cronbach's alpha measures a scale's internal consistency and reliability. Alpha for the Vocational/Educational scale in the composite norm set is 0.671. Table 145 shows the correlation of each item with the overall scale score, and the value of alpha when the item is removed from the scale. Removing *high_school* or *need_training* would increase alpha modestly, but these items are critical for covering the educational aspect of the Vocational/Educational scale, and they are well-correlated with the scale score. Removing any of the other items would lower alpha.

Scale distributions in the norm groups

Table 146: Tabulation by population and gender with mean scores. Note: this table does not present the scale score from the single transgender person in the data.

Group	N	Mean
Jail Female	405	5.3
Prison Female	1291	6.2
Probation Female	2852	5.8
Jail Male	1900	3.9
Prison Male	6124	5.4
Probation Male	8541	5.1
ALL	21113	5.2

Table 146 shows the mean Vocational/Educational scale score for each population subgroup in the normative data. These indicate show that education and employment needs among Women are more dire than among Men, and that those in prison have the greatest need among the various corrections populations.

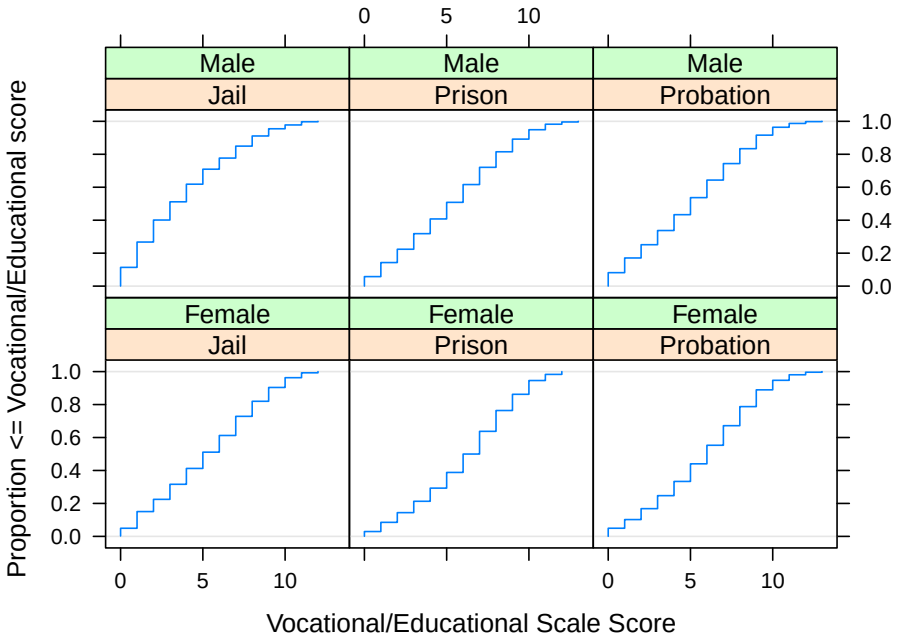


Figure 25: Cumulative distribution plots comparing subgroups in the norm data on the Vocational/Educational scale

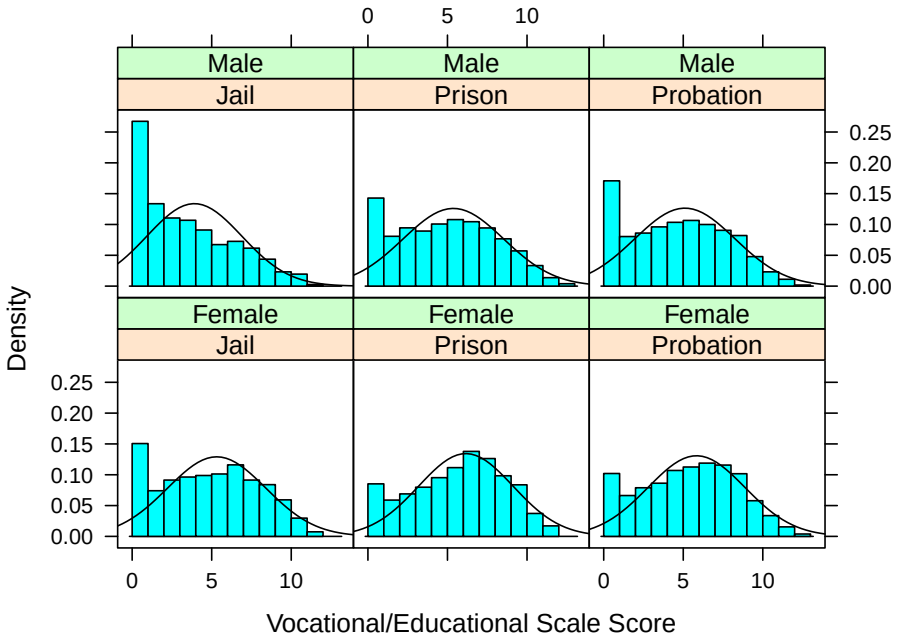


Figure 26: Histograms comparing subgroups in the norm data on the Vocational/Educational scale

The cumulative distributions for the subgroups are shown in Figure 25 and the histograms are shown in Figure 26. These plots enhance the information in Table 146. For example, a substantial

Table 147: Score ranges for the Vocational/Educational scale.

	Level		
	Unlikely	Probable	Highly Probable
GN Composite	0 – 5	6 – 7	8 – 13
GN Prison	0 – 6	7	8 – 13
GN Probation	0 – 5	6 – 7	8 – 13
GN Jail (Client-specific)	0 – 4	5 – 6	7 – 13
Male Composite	0 – 5	6 – 7	8 – 13
Male Prison	0 – 5	6 – 7	8 – 13
Male Probation	0 – 5	6 – 7	8 – 13
Male Jail (Client-specific)	0 – 3	4 – 5	6 – 13
Female Composite	0 – 6	7 – 8	9 – 13
Female Prison	0 – 7	8	9 – 13
Female Probation	0 – 6	7 – 8	9 – 13
Female Jail (Client-specific)	0 – 5	6 – 7	8 – 13

proportion of the Men in Jail score 0 or very low on this scale, contributing to their relatively low mean score on this item.

Score ranges

Table 147 displays the score ranges for each of the subgroups in the norm data. GN is a subgroup that includes all the Men and Women. Since there are more Men in the data, the GN ranges more strongly reflect the ranges for Men than for Women. Most agencies choose to use ranges based on gender-specific norms; GN is not available in the software at this time.

The ranges are determined by splitting the scores of each subgroup into the lowest 50% of scores for Unlikely, the highest 30% of scores for Highly Probable, and the remaining 20% in between for Probable. This table of ranges reflects what was seen earlier in the score distributions by subgroup.

Mokken Scale Analysis (MSA): Checking the observable measurement properties of the Monotone Homogeneity Model

Item-pair scalability coefficient The final Vocational/Educational scale variables are listed in Table 142. Table 148 displays the item-pair scalability coefficients for all Vocational/Educational scale items. All values are all positive. The smallest scalability coefficient (0.067) is between *high_school and need_training*. The largest scalability coefficient is 0.841 between *job and haveempschool*.

Table 148: Item-pair scalability coefficients (H_{ij}). Higher H_{ij} result in stronger scales; $-\infty < H_{ij} \leq 1$. The monotone homogeneity (IRT) model assumption requires that $0 \leq H_{ij} \leq 1$.

	high_school	job	skill	job_last_year	need_training	wages_above_min	chance_success_work	haveempschool
high_school		0.177	0.169	0.198	0.067	0.168	0.185	0.171
job	0.177		0.447	0.689	0.099	0.368	0.419	0.841
skill	0.169	0.447		0.489	0.209	0.348	0.343	0.411
job_last_year	0.198	0.689	0.489		0.148	0.363	0.476	0.699
need_training	0.067	0.099	0.209	0.148		0.279	0.099	0.097
wages_above_min	0.168	0.368	0.348	0.363	0.279		0.380	0.342
chance_success_work	0.185	0.419	0.343	0.476	0.099	0.380		0.392
haveempschool	0.171	0.841	0.411	0.699	0.097	0.342	0.392	

* The standard errors (se) have been removed to allow all item names to fit in the table. The standard errors range from 0.005 to 0.014. The smallest item-pair scalability coefficient is 0.067 and the largest is 0.841.

Item-scalability coefficient The item scalability coefficients are in Table 149; not all are above the lower bound of 0.3.

Table 149: Mokken Analysis: Item H values should be above 0.3

	Item H	se
high_school	0.167	(0.006)
job	0.477	(0.004)
skill	0.356	(0.005)
job_last_year	0.431	(0.005)
need_training	0.157	(0.006)
wages_above_min	0.329	(0.005)
chance_success_work	0.348	(0.007)
haveempschool	0.462	(0.004)

Test Scalability Coefficient The test scalability coefficient (H) for the Vocational/Educational scale in its final form is 0.351 which is within the limits for the monotone homogeneity model. A scalability coefficient of $H = 0.351$ implies that the scale is weak. The standard error for the statistic is (0.003). The original data had a test scalability coefficient of 0.258 with a standard error of (0.003).

The Vocational/Educational scale is actually a higher order scale: multidimensional IRT was not performed. Although the Vocational/Educational scale does not meet all Mokken scale assumptions as it is considered a higher order scale, and we would not expect it to, we do still test the assumptions of the Monotone Homogeneity Model. As with the other unidimensional scales, we test for DIF first.

Differential Item Functioning (DIF) Differential Item Functioning (DIF) was tested on a subset of the data ($n = 1000$) after variable transformations for race; the *Other* category for race

Table 150: Percent of Black, Hispanic, and White persons in a random sample of 1,000 cases drawn from the composite norm data; the data includes all variable transformations.

Ethnicity	Percent
Black	40
Hispanic	21
White	40

was not included since it only made up 7.5% of the final sample. Table 150 denotes the ethnicity proportions from the random subset of $n = 1000$ from the transformed data.

The following items were found to have differential item functioning (DIF) in the $n = 1{,}000$ subsample: ***high_school, grades_hs, skill, job_last_year, wages_above_min and haveempschool***. In the original data, a random subsample of $n = 1{,}000$ was also selected; the following variables were flagged with DIF: ***high_school, expelled, skill, need_training and wages_above_min***.

Non-Parametric Monotone Homogeneity (IRT) Model

The three assumptions for the monotone homogeneity model are the *Unidimensionality*, *Local Independence*, and *Latent Monotonicity*.

Unidimensionality The Unidimensionality assumption is tested using Principal Components Analysis (PCA). This was done in section 7.13. We see in Table 144 that the first eigenvalue (2.131) is more than twice as large as the second eigenvalue (0.818) and over half (51.8%) of the variance is explained within the first principal component. This is strong evidence that the Vocational/Educational scale is unidimensional. This is somewhat surprising since the Vocational/Educational scale is higher order, but we continue with our analysis.

Local Independence The assumption of local independence was tested using the Chi-Squared Test of Independence in R with the parameter `simulate.p.value` set to `TRUE`. The $n = 1{,}000$ subset used for testing DIF from the transformed data (items changed for the COMPAS-R Core scales) was used. There was 1 pair of items with p-values greater than or equal to 0.05. This item pairs is listed in Table 151.

Table 151: Item pair and its corresponding p-value for items that suggest independence from a subset of n = 1,000 from the final COMPAS-R CORE scale.

item1	item2	p_value
high_school	need_training	0.106

Latent Monotonicity The latent monotonicity assumption is investigated by examining the observable property of the test data known as manifest monotonicity. We see in Table 152 that there are no violations (#vi) for any of the items in the Vocational/Educational scale.

Table 152: Manifest monotonicity results: used to assess latent monotonicity. The function `check.mononotinicity` was used from the R package *mokken*.

	ItemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
high_school	0.17	15	0	0	0	0	0	0	0	0
job	0.48	21	0	0	0	0	0	0	0	0
skill	0.36	21	0	0	0	0	0	0	0	0
job_last_year	0.43	63	0	0	0	0	0	0	0	0
need_training	0.16	15	0	0	0	0	0	0	0	0
wages_above_min	0.33	84	0	0	0	0	0	0	0	0
chance_success_work	0.35	42	0	0	0	0	0	0	0	0
haveempschool	0.46	21	0	0	0	0	0	0	0	0

7.14 Residential Stability

The Residential Stability scale is derived from the Residential Instability scale in the standard COMPAS Core. In the development sample, there were 14,568 individuals with completed Residential Instability scales.

Review of Theoretical and Substantive Validity

Difficulty with stable housing is a non-criminogenic need and an example of a stabilization factor that, when addressed prior to treatment for identified criminogenic needs, enhances responsivity to that treatment (Marlowe 2018). The Residential Stability scale focuses on these difficulties, such as maintaining local ties, keeping a stable address, and other issues related to housing needs.

Scale Items

The Residential Stability scale is constructed using six items from the Social Exclusion domain of the COMPAS-R Core. These items and their short descriptions are shown in Table 153.

The Residential Stability scale items focus on whether the person has a stable and verifiable address, long-term local ties, and the stability of their residential circumstances. A high-score indicates a person who has no regular living situation, has lived at the present address for short time, has minimal local ties, and frequently changes residences.

Table 153: COMPAS-R Core Residential Stability Scale Items. The interview is instructed to ask the respondent to think of their current living situation for this scale.

Item	Short Description (Response Categories)
has_address_R	Could you provide a verifiable address? *(No=1; Yes=0)*
l_drift_R	Do you have a regular living situation (an address where you usually stay and can be reached)? *(No=1; Yes=0)*
l_situation_R	What is your living situation? *(Live with family=0; Live with friends and family=1; Live with friends=2; Live alone=2; Not with family, or friends, or alone=2)*
mnth_local_R	How long have you lived or worked in this area? *(Less than a year=1; A year or longer=0)*
res_moves_R	How often have you moved in the last 12 months? *(None=0; Once=1; Twice=2; Three times=3; Four times=4; Five or more times=5)*
yrs_address_R	How long have you lived at your current address? *(0-5 months=4; 6-11 months=3; 1-3 years=2; 4-5 years=1; 6 years or longer=0)*

Item changes from the Residential Instability scale in the standard COMPAS Core to the Residential Stability scale in the COMPAS-R Core

Two items from the Residential Instability scale were not used in the Residential Stability scale: the item that asks about frequency of contact with one's family and the item that asks whether there is

a telephone (or cell phone) at the person's residence. Removal of these two items did not affect alpha.

Additionally, three items were grouped into a single item. These were

- Do you live with family–natural parents, primary person who raised you, blood relative, spouse, children, or boyfriend/girlfriend if living together for more than 1 year?
- Do you live with friends?
- Do you live alone?

These items were found to create some contradictions within the data, while perhaps inflating the alpha reliability, since a greater number of items can contribute to a larger value of alpha. The item they were grouped into, *l_situation*, allows the collection of more consistent responses to those three items.

Overall, these changes reduced the reliability of the scale from 0.676 to 0.643, but it may be a more accurate estimate of alpha. Additional statistics for the scale are included in this section.

Factor structure in the norm data

Table 154: Principal component loadings.

	PC1	PC2	PC3
has_address	-0.032	0.059	0.001
l_drift	-0.047	0.091	0.017
mnth_local	-0.179	-0.047	-0.057
res_moves	-0.505	0.510	0.689
yrs_address	-0.825	-0.457	-0.274
l_situation	-0.169	0.719	-0.668

Table 155: Eigenvalues and the proportion of the variance explained by each principal component.

	PC1	PC2	PC3
Eigenvalue	3.022	0.752	0.692
Standard deviation	1.738	0.867	0.832
Proportion of variance	0.641	0.160	0.147
Cumulative proportion	0.641	0.801	0.947

Principal components analysis of the Residential Stability items in the COMPAS-R composite norm group is presented here. The loadings of the first three principal components are shown in Table 154. Table 155 shows the eigenvalues of the first three principal components and the proportion of the variance explained by each. As shown in the table, the first principal component has an eigenvalue of 3.022 and explains 64.1% of the variance in the data. Only three principal components are required to explain at least 90% of the variation in the data.

Scale consistency and reliability in the norm data

Table 156: Correlation of items with scale score, and alpha reliability of the scale when each item is removed. Compare with alpha of 0.643 for all six items.

	Correlation with score	Alpha when item removed
has_address	0.369	0.647
l_drift	0.430	0.638
mnth_local	0.671	0.578
res_moves	0.790	0.503
yrs_address	0.857	0.522
l_situation	0.529	0.633

Cronbach's alpha measures a scale's internal consistency and reliability. Alpha for the Residential Stability scale in the composite norm set is 0.643. Table 156 shows the correlation of each item with the overall scale score, and the value of alpha when the item is removed from the scale. Removing *has_address* affects alpha minimally, but it is an important item to have information about, and it is well-correlated with the scale score. Removing any of the other items would lower alpha.

Scale distributions in the norm groups

Table 157: Tabulation by population and gender with mean scores. Note: this table does not present the scale score from the single transgender person in the data.

Group	N	Mean
Jail Female	404	5.0
Prison Female	902	4.6
Probation Female	1613	4.6
Jail Male	1900	3.8
Prison Male	4599	4.3
Probation Male	5149	4.2
ALL	14567	4.3

Table 157 shows the mean Residential Stability scale score for each population subgroup in the normative data. The mean score for Women is higher compared to Men, regardless of corrections setting; the mean score for Men is between 3.8 and 4.3. Among the Women's scores, the mean score for Women in Jail is higher than that of either Women on Probation or in Prison.

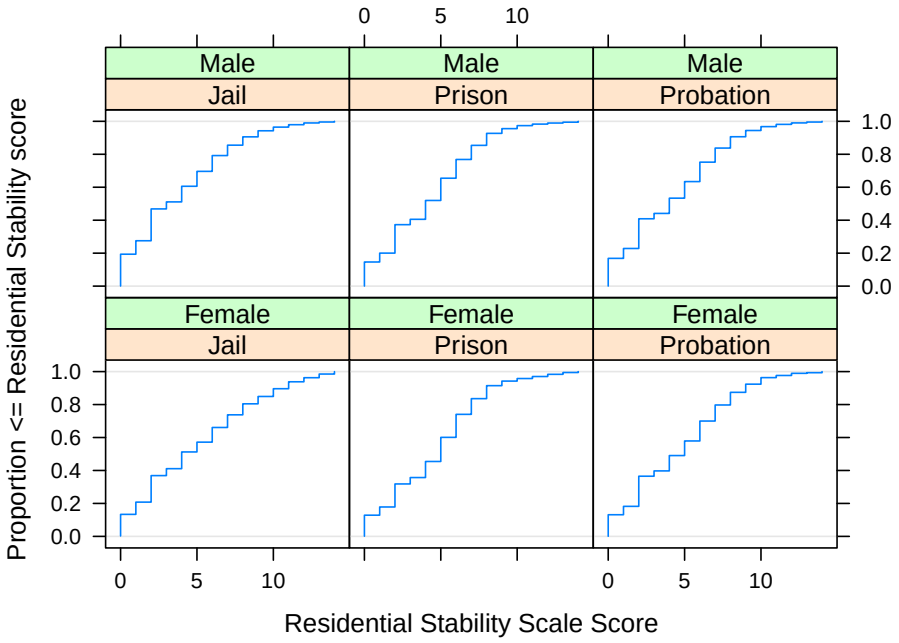


Figure 27: Cumulative distribution plots comparing subgroups in the norm data on the Residential Stability scale

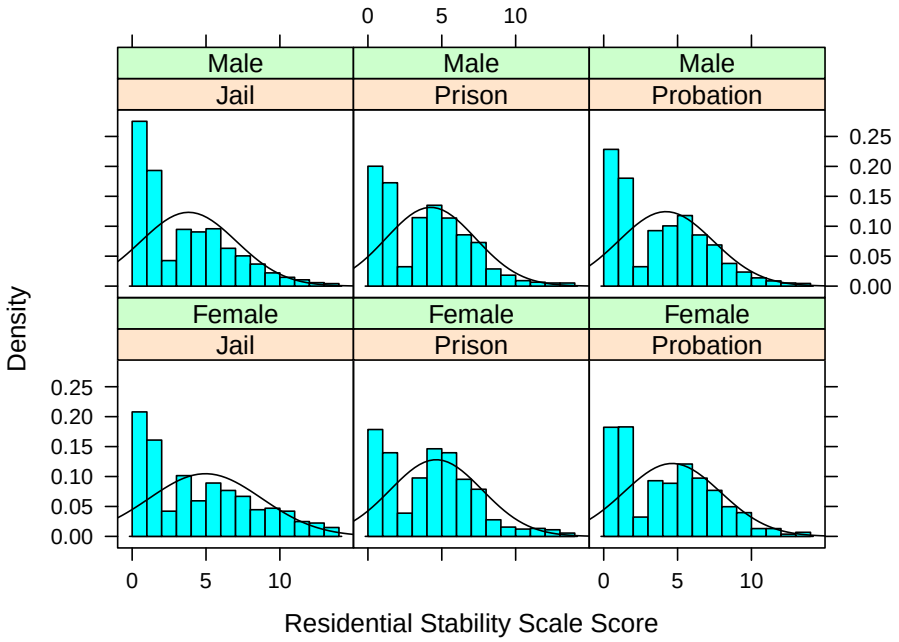


Figure 28: Histograms comparing subgroups in the norm data on the Residential Stability scale

The cumulative distributions for the subgroups are shown in Figure 27 and histograms for the subgroups are shown in Figure 28. These plots enhance the information in Table 157. All subgroups

Table 158: Score ranges for the Residential Stability scale.

	Level		
	Unlikely	Probable	Highly Probable
GN Composite	0 – 4	5 – 6	7 – 14
GN Prison	0 – 4	5 – 6	7 – 14
GN Probation	0 – 4	5 – 6	7 – 14
GN Jail (Client-specific)	0 – 4	5 – 6	7 – 14
Male Composite	0 – 4	5 – 6	7 – 14
Male Prison	0 – 4	5 – 6	7 – 14
Male Probation	0 – 4	5 – 6	7 – 14
Male Jail (Client-specific)	0 – 3	4 – 6	7 – 14
Female Composite	0 – 5	6	7 – 14
Female Prison	0 – 5	6	7 – 14
Female Probation	0 – 4	5 – 6	7 – 14
Female Jail (Client-specific)	0 – 4	5 – 7	8 – 14

contain a large proportion of very low scores, with an apparently separate group that follows a more bell-shaped distribution throughout the entire range of scores. Men in Jail exhibit the largest proportion of very low scores.

Score ranges

Table 158 shows the score ranges for each of the subgroups in the norm data. GN is a subgroup that includes all the Men and Women. Since there are more Men in the data, the GN ranges more strongly reflect the ranges for Men than for Women. Most agencies choose to use ranges based on gender-specific norms; GN is not available in the software at this time.

The ranges are determined by splitting the scores of each subgroup into the lowest 50% of scores for Unlikely, the highest 30% of scores for Highly Probable, and the remaining 20% in between for Probable. This table of ranges reflects what was seen earlier in the score distributions by subgroup.

Mokken Scale Analysis (MSA): Checking the observable measurement properties of the Monotone Homogeneity Model

Item-pair scalability coefficient The final Residential Stability scale variables are listed in Table 153. Table 159 displays the item-pair scalability coefficients for all Residential Stability scale items. All values are all positive. The smallest scalability coefficient (0.234) is between *l_situation and mnth_local*. The largest scalability coefficient is 0.858 between *yrs_address and res_moves*.

Table 159: Item-pair scalability coefficients (H_{ij}). Higher H_{ij} result in stronger scales; $-\infty < H_{ij} \leq 1$. The monotone homogeneity (IRT) model assumption requires that $0 \leq H_{ij} \leq 1$.

	has_address	se	yrs_address	se	l_drift	se	l_situation	se	mnth_local	se	res_moves	se
has_address			0.661	(0.023)	0.638	(0.018)	0.544	(0.021)	0.432	(0.025)	0.349	(0.017)
yrs_address	0.661	(0.023)			0.621	(0.020)	0.332	(0.011)	0.801	(0.005)	0.858	(0.004)
l_drift	0.638	(0.018)	0.621	(0.020)			0.502	(0.018)	0.374	(0.021)	0.430	(0.014)
l_situation	0.544	(0.021)	0.332	(0.011)	0.502	(0.018)			0.234	(0.010)	0.276	(0.011)
mnth_local	0.432	(0.025)	0.801	(0.005)	0.374	(0.021)	0.234	(0.010)			0.550	(0.010)
res_moves	0.349	(0.017)	0.858	(0.004)	0.430	(0.014)	0.276	(0.011)	0.550	(0.010)		

* The standard errors range from 0.004 to 0.025. The smallest item-pair scalability coefficient is 0.234 and the largest is 0.858.

Item-scalability coefficient The item scalability coefficients are in Table 160; they are all above the lower bound of 0.3.

Table 160: Mokken Analysis: Item H values should be above 0.3

	Item H	se
has_address	0.486	(0.013)
yrs_address	0.658	(0.005)
l_drift	0.499	(0.012)
l_situation	0.313	(0.009)
mnth_local	0.556	(0.006)
res_moves	0.572	(0.005)

Test Scalability Coefficient The test scalability coefficient (H) for the Residential Stability scale in its final form is 0.528 which is within the limits for the monotone homogeneity model. A scalability coefficient of $H = 0.528$ implies that the scale is very strong. The standard error for the statistic is (0.005). The original data had a test scalability coefficient of 0.389 with a standard error of (0.004).

Differential Item Functioning (DIF) Differential Item Functioning (DIF) was tested on a subset of the data ($n = 1{,}000$) for race, after variable transformations; the *Other* category for race was not included since it only made up 7.2% of the final sample. Table 161 denotes the ethnicity proportions from the random subset of $n = 1{,}000$ from the transformed data.

No items were found to have differential item functioning (DIF) in the $n = 1{,}000$ subsample. In the original data, a subsample of $n = 1{,}000$ was also selected; The following items were flagged for DIF: *fam_freq, l_fam and l_friends*. The item *fam_freq* was not included in the final COMPAS-R Core scale, *Residentail Stability*; The other two items were combined with *l_alone* to create the variable *l_situation*.

Table 161: Percent of Black, Hispanic, and White persons in a random sample of 1,000 cases drawn from the composite norm data; the data includes all variable transformations.

Ethnicity	Percent
Black	40
Hispanic	17
White	43

Non-Parametric Monotone Homogeneity (IRT) Model

The three assumptions for the monotone homogeneity model are the *Unidimensionality*, *Local Independence*, and *Latent Monotonicity*.

Unidimensionality The Unidimensionality assumption is tested using Principal Components Analysis (PCA). This was done in section 7.14. We see in Table 155 that the first eigenvalue (3.022) is much larger than the remaining two and almost twice as large as the second eigenvalue (1.738). Furthermore, 64.1% of the variance is explained within the first eigenvalue. This is strong evidence that the Residential Stability scale is unidimensional.

Local Independence The assumption of local independence was tested using the Chi-Squared Test of Independence in R with the parameter `simulate.p.value` set to `TRUE`. All p-values were found to be less than 0.001, indicating evidence of local dependence.

Latent Monotonicity The latent monotonicity assumption is investigated by examining the observable property of the test data known as manifest monotonicity. We see in Table 162 that there are three violations (#vi) for the item ***`yrs_address`***.

Table 162: Manifest monotonicity results: used to assess latent monotonicity. The function `check.mononotinicity` was used from the R package *mokken*.

	ItemH	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
has_address	0.49	15	0	0.00	0.00	0.00	0.0000	0.00	0	0
yrs_address	0.66	40	3	0.07	0.15	0.27	0.0068	11.26	3	99
l_drift	0.50	15	0	0.00	0.00	0.00	0.0000	0.00	0	0
l_situation	0.31	20	0	0.00	0.00	0.00	0.0000	0.00	0	0
mnth_local	0.56	15	0	0.00	0.00	0.00	0.0000	0.00	0	0
res_moves	0.57	47	0	0.00	0.00	0.00	0.0000	0.00	0	0

7.15 Cognitive Behavioral*

The Cognitive Behavioral Scale is derived from the Cognitive Behavioral scale in the standard COMPAS Core. It combines four of the COMPAS-R Core scales: *Associates and Peers*; *Society and Routines*; *Thinking and Attitudes*; and *Socialization History*. It measures aspects of Family, Antisocial Personality Pattern, and Procriminal Associates, which are all criminogenic needs within the Central Eight. It is considered a higher order scale and most of the analyses that have been done on the scales that make up the Cognitive Behavioral scale do not make sense to deploy here. We do provide some analytical results for this scale below.

Scale Items

There are 35 items that make up the Cognitive Behavioral needs scale. The *Associates and Peers* scale is made up of 5 items: four items have points that range from 0 - 3 and one item has points that range from 0 - 1 as discussed in section 7.5 and displayed in Table 51. The item ***friends_drugs*** from the *Associates and Peers* scale also belongs to the *Society and Routines* scale, composed of 12 items. This shared item maintains the same point system regardless of which scale it belongs to and is thus subtracted from the total Cognitive Behavioral scale when the four scale scores are summed together to calculate the Cognitive Behavioral needs score. The *Thinking and Attitudes* scale is composed of 9 items, where all 9 items range in points between zero and three, contributing anywhere from 0 to 27 total points to the *Cognitive Behavioral scale*. The *Socialization History* scale, the last of the four scales that make up the Cognitive Behavioral scale, is composed of 10 items. The scale contributions (possible **Points**) for each scale can be seen in Table 163. The question text for the shared item, ***friends_drugs***, belonging to both the Associates and Peers and Society and Routines scales is displayed in the last row, where it is subtracted.

Table 163: Cognitive Behavioral Scale (Standard Version)

Points	Scale or item
0 - 13	Associates and Peers
0 - 21	Society and Routines
0 - 27	Thinking and Attitudes
0 - 22	Socialization History
- (0 - 3)	How many of your friends/acquaintances are taking illegal drugs regularly (more than a couple times a month)?

Table 164: Cognitive Behavioral Scale (Conviction Version)

Points	Scale or item
0 - 13	Associates and Peers
0 - 21	Society and Routines
0 - 27	Thinking and Attitudes
0 - 22	Socialization History (Conviction Version)
- (0 - 3)	How many of your friends/acquaintances are taking illegal drugs regularly (more than a couple times a month)?

Table 164 displays the scale inputs for the Cognitive Behavioral Conviction Version. The difference

between the non-conviction and conviction version is the *Socialization History Conviction Version* scale. This table is presented for documentation purposes. Analysis regarding this scale was not performed. The remaining analyses below are only for the non-conviction version of the Cognitive Behavioral scale.

Item changes from the Cognitive Behavioral scale in the standard COMPAS Core

The Cognitive Behavioral scale in the standard COMPAS Core is composed of five need scales:

1. Criminal Associates and Peers: Associates and Peers in COMPAS-R Core
2. Criminal Opportunity: Society and Routines in COMPAS-R Core
3. Criminal Thinking Self-Report: Thinking and Attitudes in COMPAS-R Core
4. Socialization Failure: Socialization History in COMPAS-R Core
5. Social Adjustment Problems: Not included in COMPAS-R Core

All changes made for the individualized scales in COMPAS-R Core were incorporated into all inputs for the COMPAS-R Core Cognitive Behavioral scale. In depth scale transformations and analyses for the five scales listed above can be found in their respective sections.

Scale consistency and reliability in the norm data

Cronbach's alpha measures a scale's internal consistency and reliability. Alpha for the Cognitive Behavioral scale in the composite norm set is 0.826. Table 165 shows the correlation of each item with the overall scale score, and the value of alpha when the item is removed from the scale.

Table 165: Correlation of items with scale score, and alpha reliability of the scale when each item is removed. Compare with alpha of 0.826 for all thirty-five items.

	Correlation with score	Alpha when item removed
Associates & Peers		
friends_arrest	0.551	0.816
friends_drugs	0.545	0.817
friends_gang	0.447	0.820
friends_jail	0.559	0.816
gang_member	0.249	0.826
Society & Routines		
res_moves	0.335	0.829
l_drift	0.229	0.825
job	0.270	0.825
job_last_year	0.331	0.835
local_crime	0.307	0.824
local_gangs	0.326	0.824
citizens_weapons	0.319	0.824
nothing	0.396	0.822
restless	0.364	0.823
need_training	0.296	0.824
haveempschool	0.287	0.824
Thinking & Attitudes		
hungry_steal	0.476	0.819
law_doesnt	0.477	0.819
law_nochance	0.454	0.820
trouble_no	0.458	0.820
dirt	0.436	0.820
roughly	0.516	0.818
insults	0.487	0.819
threaten	0.541	0.817
stolen_rich	0.489	0.819
Socialization History		
n_juv_felony	0.302	0.827
juv_placement	0.287	0.824
high_school	0.225	0.826
expelled	0.404	0.822
conflict_w_teacher	0.460	0.820
skipped_classes	0.455	0.822
fights_at_school	0.448	0.820
parent_jailed	0.335	0.823
parent_drug_problem	0.298	0.824
conv_parent	0.356	0.823

* The item friends_drugs is in both the Associates & Peers and Society & Routines scales; however, to avoid redundancy, it is only listed under the Associates & Peers scale in the above table.

Table 166: Score ranges for the Cognitive Behavioral scale

	Level		
	Unlikely	Probable	Highly Probable
GN Composite	0 - 15	16 - 21	22 - 80
GN Prison	0 - 16	17 - 23	24 - 80
GN Probation	0 - 14	15 - 21	22 - 80
GN Jail (Client-specific)	0 - 17	18 - 24	25 - 80
Male Composite	0 - 15	16 - 22	23 - 80
Male Prison	0 - 16	17 - 23	24 - 80
Male Probation	0 - 15	16 - 21	22 - 80
Male Jail (Client-specific)	0 - 17	18 - 24	25 - 80
Female Composite	0 - 14	15 - 21	22 - 80
Female Prison	0 - 15	16 - 22	23 - 80
Female Probation	0 - 14	15 - 20	21 - 80
Female Jail (Client-specific)	0 - 19	20 - 27	28 - 80

Score ranges

Table 166 displays the score ranges for each of the subgroups used in the norm data. The subgroup GN contains both men and women, but predominately men since they represent the majority of the sample; thus the range values strongly reflect the ranges for Men. GN is currently not available in the software for agency use. Table 167 displays the score ranges for each of the subgroups used in the conviction norm data.

The ranges are determined for the Cognitive Behavioral scale by splitting the scores of each subgroup into the lowest 40% of scores for Unlikely, the highest 30% of scores for Highly Probable and the remaining 30% in between for Probable.

Mokken Scale Analysis (MSA): Checking the observable measurement properties of the Monotone Homogeneity Model

Item-pair scalability coefficient Table 168 and the tables that follow, display the item-pair scalability coefficients for most of Cognitive Behavioral scale items. All values are not positive. The smallest scalability coefficient (-0.051) is between *stolen_rich and parent_drug_problem*. The largest scalability coefficient is 0.881 between *parent_jailed and conv_parent*. Since there are thirty-five items in the Cognitive Behavioral scale, the item-pair scalability coefficient table was broken up. Not all item pairs are shown. However, the tables that contain the smallest and largest item-pair scalability coefficients are displayed below.

The presence of negative item-pair scalability coefficients is a violation of the Mokken scale assumptions as well as the Monotone Homogeneity Model (MHM). Thus, no other MSA was done for this scale: neither IRT analysis for the MHM assumptions, nor dfferential item functioning (DIF). However, DIF was completed on all subscales that make up the Cognitive Behavioral scale in COMPAS-R Core. (See the corresponding sections for details.)

Further CTT, IRT, or mokken scale analysis was not attempted. The scale is considered a higher order scale; it combines four other COMPAS-R Core scales where CTT and IRT analyses were

Table 167: Score ranges for the Cognitive Behavioral (Conviction Version) scale. Scores and cut points for this scale were determined using scores from the four COMPAS-R Core scales that make up the Cognitive Behavioral (Conviction Version) scale: *Criminal Associates and Peers*; *Society and Routines*; *Thinking and Attitudes*; and *Socialization History*.

	Level		
	Unlikely	Probable	Highly Probable
GN Composite	0 - 15	16 - 21	22 - 80
GN Prison	0 - 16	17 - 22	23 - 80
GN Probation	0 - 14	15 - 20	21 - 80
Male Composite	0 - 15	16 - 21	22 - 80
Male Prison	0 - 16	17 - 22	23 - 80
Male Probation	0 - 14	15 - 20	21 - 80
Female Composite	0 - 13	14 - 19	20 - 80
Female Prison	0 - 14	15 - 20	21 - 80
Female Probation	0 - 13	14 - 19	20 - 80

Table 168: Item-pair scalability coefficients (H_{ij}). Higher H_{ij} result in stronger scales; $-\infty < H_{ij} \leq 1$. The monotone homogeneity (IRT) model assumption requires that $0 \leq H_{ij} \leq 1$.

	friends_arrest	friends_drugs	friends_gang	friends_jail	gang_member	res_moves	l_drift	job	job_last_year
friends_arrest		0.642	0.515	0.876	0.287	0.175	0.175	0.067	0.110
friends_drugs	0.642		0.489	0.622	0.298	0.193	0.198	0.012	0.033
friends_gang	0.515	0.489		0.515	0.472	0.107	0.129	0.055	0.223
friends_jail	0.876	0.622	0.515		0.250	0.176	0.159	0.078	0.126
gang_member	0.287	0.298	0.472	0.250		0.076	0.082	-0.049	0.137
res_moves	0.175	0.193	0.107	0.176	0.076		0.430	0.081	0.192
l_drift	0.175	0.198	0.129	0.159	0.082	0.430		0.213	0.328
job	0.067	0.012	0.055	0.078	-0.049	0.081	0.213		0.716
job_last_year	0.110	0.033	0.223	0.126	0.137	0.192	0.328	0.716	

* The standard errors (se) have been removed to allow all item names to fit in the table. The standard errors range from 0.008 to 0.085. The smallest item-pair scalability coefficient is -0.051 and the largest is 0.881.

	local_crime	local_gangs	citizens_weapons	nothing	restless	need_training	haveempschool	hungry_steal	law_doesnt
friends_arrest	0.215	0.188	0.226	0.232	0.244	0.190	0.077	0.157	0.103
friends_drugs	0.247	0.231	0.262	0.265	0.321	0.232	0.040	0.231	0.175
friends_gang	0.223	0.298	0.267	0.161	0.177	0.325	0.070	0.213	0.162
friends_jail	0.197	0.179	0.202	0.220	0.252	0.177	0.081	0.175	0.131
gang_member	0.222	0.293	0.256	0.097	0.179	0.352	0.007	0.167	0.193
res_moves	0.138	0.096	0.124	0.148	0.172	0.144	0.087	0.111	0.074
l_drift	0.121	0.049	0.089	0.160	0.197	0.181	0.254	0.149	0.102
job	0.044	-0.011	0.034	0.169	0.135	0.105	0.869	0.027	0.019
job_last_year	0.115	0.098	0.117	0.221	0.146	0.164	0.743	0.002	-0.024

performed. Links for each scale analysis are included above in the section: Item changes from the Cognitive Behavioral scale in the standard COMPAS Core. This scale was not expected to meet any of the mokken scale analysis and the analyses that were performed support that expectation.

	law_nochance	trouble_no	dirt	roughly	insults	threaten	stolen_rich	n_juv_felony	juv_placement
friends_arrest	0.094	0.132	0.098	0.148	0.179	0.173	0.116	0.122	0.104
friends_drugs	0.117	0.152	0.085	0.187	0.210	0.242	0.156	0.104	0.102
friends_gang	0.085	0.121	0.137	0.246	0.255	0.256	0.209	0.215	0.177
friends_jail	0.116	0.144	0.091	0.164	0.201	0.191	0.140	0.121	0.109
gang_member	0.132	0.202	0.154	0.192	0.241	0.275	0.213	0.221	0.193
res_moves	0.051	0.060	0.016	0.044	0.093	0.075	0.059	0.030	0.048
l_drift	0.085	0.091	0.078	0.102	0.113	0.094	0.120	0.031	0.045
job	0.090	0.042	0.062	0.052	-0.002	0.036	0.072	0.033	0.079
job_last_year	0.055	0.019	0.032	0.012	-0.017	0.001	0.019	0.228	0.295

	law_nochance	trouble_no	dirt	roughly	insults	threaten	stolen_rich	n_juv_felony	juv_placement
law_nochance		0.595	0.583	0.599	0.531	0.448	0.628	-0.002	0.037
trouble_no	0.595		0.610	0.585	0.457	0.435	0.598	-0.008	0.013
dirt	0.583	0.610		0.703	0.519	0.636	0.659	0.049	0.089
roughly	0.599	0.585	0.703		0.660	0.763	0.648	0.047	0.087
insults	0.531	0.457	0.519	0.660		0.689	0.623	0.058	0.072
threaten	0.448	0.435	0.636	0.763	0.689		0.582	0.083	0.099
stolen_rich	0.628	0.598	0.659	0.648	0.623	0.582		0.074	0.125
n_juv_felony	-0.002	-0.008	0.049	0.047	0.058	0.083	0.074		0.617
juv_placement	0.037	0.013	0.089	0.087	0.072	0.099	0.125	0.617	

	expelled	conflict_w_teacher	skipped_classes	fights_at_school	parent_jailed	parent_drug_problem	conv_parent
law_nochance	0.109	0.111	0.083	0.136	0.012	0.017	0.025
trouble_no	0.134	0.148	0.075	0.140	0.057	0.031	0.048
dirt	0.091	0.154	0.078	0.163	0.031	-0.037	0.004
roughly	0.127	0.179	0.139	0.235	0.073	-0.014	0.043
insults	0.204	0.216	0.132	0.227	0.112	0.052	0.105
threaten	0.202	0.254	0.139	0.301	0.112	0.043	0.074
stolen_rich	0.099	0.138	0.091	0.168	-0.000	-0.051	0.005
n_juv_felony	0.557	0.165	0.204	0.231	0.171	0.091	0.158
juv_placement	0.536	0.179	0.198	0.223	0.198	0.088	0.181

* The standard errors (se) have been removed to allow all item names to fit in the table. The standard errors range from 0.008 to 0.085. The smallest item-pair scalability coefficient is -0.051 and the largest is 0.881. The smallest can be found in this table.

	expelled	conflict_w_teacher	skipped_classes	fights_at_school	parent_jailed	parent_drug_problem	conv_parent
high_school	0.185	0.107	0.107	0.124	0.089	0.014	0.109
expelled		0.696	0.434	0.667	0.371	0.222	0.355
conflict_w_teacher	0.696		0.484	0.401	0.232	0.175	0.219
skipped_classes	0.434	0.484		0.375	0.205	0.196	0.186
fights_at_school	0.667	0.401	0.375		0.208	0.159	0.185
parent_jailed	0.371	0.232	0.205	0.208		0.423	0.881
parent_drug_problem	0.222	0.175	0.196	0.159	0.423		0.481
conv_parent	0.355	0.219	0.186	0.185	0.881	0.481	

* The standard errors (se) have been removed to allow all item names to fit in the table. The standard errors range from 0.008 to 0.085. The smallest item-pair scalability coefficient is -0.051 and the largest is 0.881. The largest can be found in this table.

7.16 Recidivism Risk: Summative GRRS

In this section we discuss the development and use of the Summative General Recidivism Risk scale. Notably omitted from the scales offered in the COMPAS-R Core is a risk scale that measures a person's potential for violent recidivism. This is because violent recidivism occurs at a very low rate and can be better predicted by considering the General Recidivism Risk Score in conjunction with the History of Violence and History of Non-compliance scores (see section 4.1.5 of the Practitioner's Guide to COMPAS Core, (Northpointe Inc. 2019), *On Counter-Intuitive Predictions*).

7.16.1 Theoretical and Substantive Validity

The COMPAS-R Summative General Recidivism Risk Scale (Summative GRRS) is based on the work of Northpointe researchers who developed the standard COMPAS General Recidivism Risk Scale (GRRS). That scale was developed to predict new offenses subsequent to the COMPAS assessment date, closely concurrent with community start. In an effort to maintain continuity for agencies who have been using the standard COMPAS assessment, the decision was made to use inputs similar to those used in the standard COMPAS GRRS, refitting the model based on the modified inputs described in this document. The inputs for the standard COMPAS GRRS were selected through diagnostic modeling strategies which are discussed briefly in the Method section below (Subsection 7.16.3). The outcome used for the original scale construction was whether a person was arrested for a new misdemeanor or felony offense within two years of the COMPAS screening date. For the COMPAS-R Summative GRRS, the outcome was whether a person was arrested for a new misdemeanor or felony offense within *three* years of the COMPAS-R screening date.

Previous sections of this chapter show the results of refining the need scales, wherein we emphasized scale reliability and coherence measures. In this section, however, the focus is on validating a risk instrument, which depends on how well recidivism is predicted. The purpose of the Summative GRRS is to predict which persons will commit crimes subsequent to their initial COMPAS screening date.

7.16.2 Scale items and scoring

The inputs used to predict recidivism are shown in Table 169, derived using the process described in Subsection 7.16.3. Please refer to Table 3 for information about the items in the Legal System Involvement Scale, Table 142 for information about the items in the Vocational/Educational Scale, and Table 170 for information about the items in the Drug Problems scale. Risk contribution development is described in detail in Appendix 10.2. A scoring sheet from the assessment is included in this section on page 138.

Summative General Recidivism Risk Scale (S–GRRS)

Legal System Involvement												
Need scale score	0	1	2	3	4	5	6 – 7	8	9 – 10	11 – 12	13 – 15	
Risk Contribution	0	4	7	8	10	11	12	13	14	15	16	13

Vocational/Educational						
Need scale score	0 – 1	2 – 5	6 – 8	9 – 12	13	
Risk Contribution	0	1	2	3	4	1

Drug Problems							
Need scale score	0	1	2	3	4	5	
Risk Contribution	0	1	2	3	4	5	2

Age at First Arrest	10 – 14	15 – 17	18 – 20	21 – 24	25 – 28	29 – 33	34 – 39	40 –46	47 – 54	55 – 64	65 – 75	76 +	
Risk Contribution	11	10	9	8	7	6	5	4	3	2	1	0	7

Assessment Age*	16 – 18	19	20 – 21	22 – 23	24 – 26	27 – 30	31 – 34	35 – 41	42 – 49	50 – 60	61 – 74	75 +	
Risk Contribution	11	10	9	8	7	6	5	4	3	2	1	0	5

*Assessment Age can be thought of as the person's age at their *nearest* birthday.
Actual calculation details are on page 173 of the COMPAS-R Core Research Report.

Arrest Rate

Number of previous arrests / Years since first arrest	0	1-2	3	4	5 – 6	7 – 8	9 – 12	13 – 18	19 +
0	10	13	16	18	19	20	21	23	24
1 – 2		10	13	14	15	16	18	19	21
3 – 4		8	11	12	13	14	15	17	18
5 – 6		6	9	10	11	13	14	16	17
7 – 9		5	8	9	10	11	13	14	16
10 – 12		4	7	8	9	10	12	13	14
13 – 15		3	6	7	8	9	11	12	14
16 – 20		2	5	6	7	8	10	11	13
21 – 26		1	4	5	6	7	9	10	12
27 +		0	3	4	5	6	7	9	10

9

Total Risk score **33**

Risk Level **Low**

Table 169: Inputs for the Summative GRRS. The scoring page indicates how values from the inputs are mapped to risk contributions. Note that age and age at first arrest are negatively associated with higher risk. That is, lower values of these inputs are associated with higher risk, and vice versa.

Input description	Range of inputs	Risk contribution range
Legal System Involvement scale score	0 – 15	0 – 16
Vocational/Educational scale score	0 – 13	0 – 4
Drug Problems scale score	0 – 5	0 – 5
Age at assessment (age at nearest birthday)	16+	0 – 11
Age at first arrest	8+	0 – 11
Arrest rate	table lookup	0 – 24
Offense or arrest opioid related	No, Yes	Not scored*
Total		0 – 71

* This item is for data collection and is not used to score the Summative GRRS. However, during scale set configuration with the Probation and Composite norms, agencies can specify that an affirmative response to this item be used to increase a person's recommended supervision level by default. The goal in making this configuration option available is to help ensure appropriate supervision for opioid-involved individuals.

Table 170 lists the items in the Drug Problems scale, whose score is used in the Summative GRRS, along with the opioid question. The items that make up this scale are only scored for the Summative GRRS; they are grouped together due to their homogeneous nature, but the Drug Problems scale score is not interpretable: this scale should not be used as a screening tool. For these reasons, this scale is not displayed on the Person Summary.

Table 170: COMPAS-R Core Drug Problems scale items

Item	Short Description (Response Categories)
ad_arrest_R	Were you using alcohol or drugs (including opioids) at the time of the current offense? *(No=0; Yes, alcohol only=0; Yes, drugs only=1; Yes, both=1)*
benefit_rx_ad_R	Do you think you would benefit from treatment for alcohol, or drugs, or both? *(No=0; Yes, alcohol only=0; Yes, drugs only=1; Yes, both=1)*
currdrg_poss_R	Is the current charge drug possession? *(Not checked in Current Charges table=0; Checked in Current Charges table=1)*
currdrg_traf_R	Is the current charge drug trafficking? *(Not checked in Current Charges table=0; Checked in Current Charges table=1)*
ever_rx_ad_R	Have you ever been in formal treatment for alcohol or drugs, such as counseling, outpatient, inpatient, or residential? *(No=0; Yes, alcohol only=0; Yes, drugs only=1; Yes, both=1)*
op_arrest_R	Is the current offense opioid related (were opioids involved at the time of offense or arrest)? *(No; Yes; Not scored*)*

* This item is for data collection and is not used to score the Summative GRRS. However, during scale set configuration with the Probation and Composite norms, agencies can specify that an affirmative response to this item be used to increase a person's recommended supervision level by default. The goal in making this configuration option available is to help ensure appropriate supervision for opioid-involved individuals.

7.16.3 Method

First, we briefly describe the method that was used to develop the *original* predictive model of recidivism for the standard COMPAS GRRS. We follow by describing the sample used to develop the Summative GRRS and methods used to transform the resulting model to a summative scale.

The first step in developing the standard COMPAS Core GRRS was to identify a candidate set of variables, ones that were correlated with new arrest and had been associated with recidivism in the criminological literature and previous research. Northpointe researchers focused on three candidate sets. Then, standard logistic regression was used to predict recidivism with the full set of variables in each candidate set. Bootstrap validation was used to estimate predictive validity and regression coefficients were modified using a penalized maximum likelihood estimation procedure. Applying this procedure to the three candidate sets resulted in three distinct models from which the final model was chosen. Selection criteria emphasized parsimony and predictive ability.

To update the COMPAS Core GRRS to the COMPAS-R Core Summative GRRS, the first step was to update the input scales: the Criminal Involvement scale (updated to the Legal System Involvement scale), the Vocational/Education scale (updated to the Vocational/Educational scale), and the Drug Problems scale. These modified inputs were used with a new norm set for the GRRS, described in Subsection 7.16.4, to fit a new logistic regression model. Appendix 10.2 describes the steps taken to transform this new model into the summative scale that is shown on page 138.

7.16.4 Sample

For this project, two large state agencies contributed recent COMPAS, supervision, and re-arrest data, allowing us to extract development and test sets; we refer to the combined data as the *risk data set*. The development sample for the risk scale differs from the normative sample in that its composition is only probationers or parolees who had three years follow time in the community after taking the COMPAS Core assessment and information on whether they experienced a new arrest within the three year period. For the development sample, the largest possible sub-sample of the risk data set was randomly chosen that provided equal numbers of Men and Women, and Blacks and Whites, with some Hispanics and members of Other races (see Table 171). In particular, Hispanics were not abundantly represented in the data, either because of the racial compositions of the populations from which they were drawn, or because the contributing agencies did not record Hispanic as a separate race, in accordance with current U.S. Census Bureau guidelines (Compton et al. 2010). Whether and how to record Hispanic as a separate race is an ongoing area of study and a matter of discretion for local justice agencies.

Table 171: Race and Gender in the Summative GRRS model development sample.

	Black	Hispanic	Other	White
Men	2667	78	87	2667
Women	2667	65	100	2667

Following model fitting using the sample shown in Table 171, the data in the norm set, described in Chapter 5 Table 1, was scored with the Summative GRRS. Score ranges were established using that data set for Low, Medium, and High text levels from the composite norm group and subgroups. See Table 173.

Finally, the test sample from the risk data set was scored using the Summative GRRS. Recidivism rates among Probationers in the test sample were calculated and are reported along with Supervision Levels for the Probation norm, Chapter 8.

To summarize, the three steps used to develop, norm, and test the model were:

1. Fit the model to the development sample ($N = 10{,}998$) from the risk data set. Transform the resulting model to a summative scale.
2. Score and cut the norm set ($N = 12{,}426$) and subsets into groups with 40% in Low, 30% in Medium, and 30% in High to find the score ranges that align with these ratios.
3. Use the test sample of Probationers ($N = 41{,}769$) from the risk data set to find the proportions failing in each level.

Correlation between COMPAS-R Summative GRRS and standard COMPAS GRRS

The COMPAS-R Summative GRRS scores have strong correlation, $r = 0.95$, with the standard COMPAS GRRS scores in the norm data. Figure 29 displays a scatter plot of scores on both scales in the norm data.

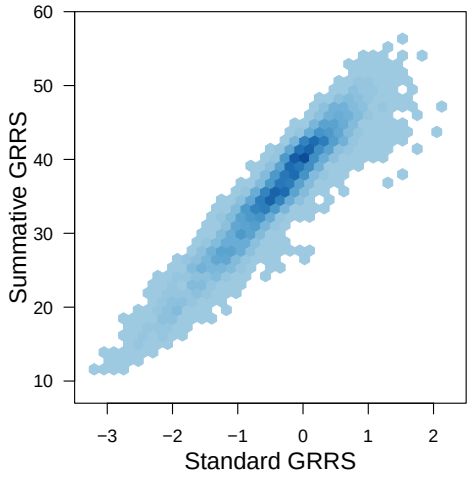


Figure 29: Scatterplot showing the correspondence between the standard GRRS and the Summative GRRS in the norm data.

Scale distributions in norm groups

Table 172: Mean Summative GRRS score in the COMPAS-R norm sub-groups.

Group	N	Mean
Jail Female	405	39.3
Prison Female	856	35.9
Probation Female	1561	33.3
Jail Male	1900	39.1
Prison Male	3350	35.9
Probation Male	4353	35.4

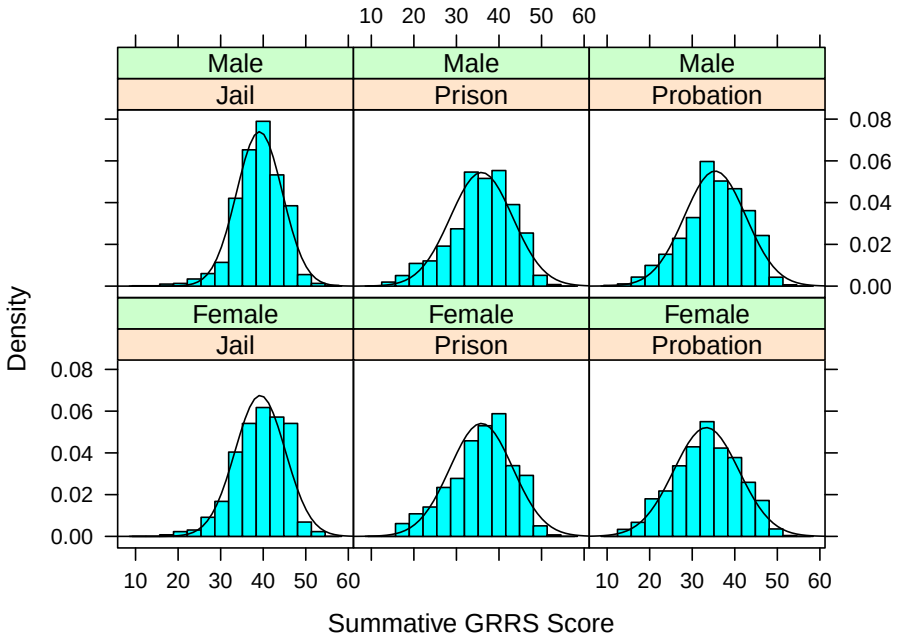


Figure 30: Histograms comparing subgroups in the norm data on the Summative GRRS

The histograms for the subgroups are shown in Figure 30. The distributions for most groups are slightly left-skewed, with the exception being the symmetrical distribution of scores from Women Probationers.

Score ranges

To establish the score ranges, scores from the norm sample were divided into the lowest 40%, the next 30%, and the highest 30% from each subgroup. These ranges are shown in Table 173.

A new question was added to the risk assessment, *Is the current offense opioid related (were opioids involved at the time of offense or arrest)?*, with the ToolTip: *Was the person under the influence of opioids at the time the current offense was committed? Answer No if the respondent was following the prescription for an opioid, as recommended by a medical professional. Answer Yes if the respondent was not following the prescription for an opioid, as recommended by a medical professional, or if they were using an opioid without a prescription.* This question is used to collect data, and is not used in scoring the Summative GRRS. However, for the both the Probation and Composite norms, agencies who wish to use an affirmative response to this item to increase a person's recommended supervision level by default, may select that option during scale set configuration in the Northpointe Suite software. The purpose of making this adjustment available is to help ensure appropriate supervision for opioid-involved individuals.

Table 173: Summative GRRS score ranges for the three risk levels.

	Level		
	Low	Medium	High
Male Composite	0 – 35	36 – 40	41 – 71
Male Prison	0 – 35	36 – 40	41 – 71
Male Probation	0 – 34	35 – 40	41 – 71
Male Jail (Client-specific)	0 – 38	39 – 42	43 – 71
Female Composite	0 – 33	34 – 40	41 – 71
Female Prison	0 – 35	36 – 40	41 – 71
Female Probation	0 – 31	32 – 38	39 – 71
Female Jail (Client-specific)	0 – 38	39 – 43	44 – 71

Criterion Validity

Criterion Validity evaluates how well a risk instrument predicts a predefined outcome, or *criterion*. For the Summative GRRS, the outcome is whether a person failed (was arrested for a new offense) within three years of an assessment administered close to their community start. We provide two standard measures of criterion validity. The first, discriminative ability, is a scale's ability to distinguish between those who eventually failed and those who did not fail in the specified time frame. The second is predictive ability. This is the more useful and appropriate measure for a risk assessment. Examining the discriminative ability of an instrument's performance is retrospective, treating recidivism as if it were a pre-existing condition. Examining predictive ability, on the other hand, is prospective, providing information at the time of assessment about a person's potential for recidivism.

Discriminative Ability: AUC/ROC Discriminative ability is measured by AUC, the area under the receiver operating characteristic curve. Larger areas under the curve indicate better performance. The area under the curve for a scale with no predictive power is .50. If the scale predicts perfectly, the area under the curve is 1. The AUC value is loosely defined as the probability that a randomly selected person who eventually recidivated has a higher risk score than a randomly selected person who did not. The conventional standard in criminal justice applications is that .65 to .69 is acceptable discrimination and .70 to .75 is good discrimination. Table 174 shows that the Summative GRRS has at least acceptable discrimination for Men and Women Probationers across race. Using the R package *pROC* (Robin et al. 2011) DeLong's test for two ROC curves was implemented to compare differences between the values. For a significance level of 0.05, the only evidence of a difference was between AUCs for White Women and Black Women (p-value = 0.020).

Table 174: AUCs for the Overall Probation test sample, as well as for Men, Women, and Blacks and Whites within gender in the Probation test sample. The only signifcant difference is between Black and White Women.

Probation test sample			AUC
Overall			0.684
	Men		0.683
		Black	0.686
		White	0.680
	Women		0.681
		Black	0.707
		White	0.675

Predictive Ability The most important measure in determining predictive ability is the Positive Predictive Value, or the PPV. It is defined as the proportion in the high risk level who fail within the predefined amount of time. This number should be higher than the base rate, preferably by 10 percentage points. Among Probationers in the test sample, the base rate for men was 0.40 and for women it was 0.36. Table 175 shows the PPVs for Men and Women Probationers, and Blacks and Whites within gender in the test sample. In each subgroup, the PPV is greater than the respective base rate by at least 20 percentage points. This indicates good predictive ability for the COMPAS-R Summative GRRS.

In addition to PPV, we also want to consider the non-recidivism rate of those who are not classified as High risk. This is done by looking at the Negative Predictive Value (NPV), defined as the proportion *not* in the high risk level who *do not* fail within the predefined amount of time. This number should be higher than 1 minus the base rate: .60 for men and .64 for women in this sample. Table 175 shows the NPVs for Men and Women Probationers, and Blacks and Whites within gender in the test sample. In each subgroup, the NPV is greater than 1 minus the respective base rate. This indicates that the people who are *not* classified as high are recidivating at a lower rate than the overall sample.

Table 175: Predictive values for Men, Women, and Blacks and Whites within gender in the Probation test sample. All positive predictive Values (PPV) are higher than the respective base rates for Men (.40) and Women (.36). All negative predictive values (NPVs) are higher than 1 minus the respective base rates for Men (.60) and Women (.64).

Probation test sample		PPV	NPV
Men		0.63	0.65
	Black	0.66	0.63
	White	0.61	0.66
Women		0.58	0.69
	Black	0.56	0.75
	White	0.59	0.67

7.17 In development: COMPAS-R Trauma Screen

The COMPAS-R Trauma Screen (in-development version) is a self-report questionnaire consisting of 24 items, including a qualifying item and all five items from the Primary Care PTSD Screen for DSM-5 (PC-PTSD-5) (Prins et al. 2016). The contents of the screen are provided in this section, along with the rationale for test item inclusion. The purpose of the COMPAS-R Trauma Screen will be to screen individuals for the likelihood of Post-Traumatic Stress Disorder (PTSD). It is not meant to diagnose PTSD, but to provide information that can help a practitioner decide if an individual is a candidate for a structured interview with a trained professional. The items from the PC-PTSD-5 are embedded among the test items of the COMPAS-R Trauma Screen (in-development version). During the development phase, a person's score from the PC-PTSD-5 will be reported. This should not be misconstrued as a score from the COMPAS-R Trauma Screen.

Qualifying question The first item in Table 176 is *traumevent_R*, which asks whether the person agrees that "An extraordinarily upsetting event occurred in my life at least one (1) month ago." This serves as a qualifying question for the remaining items. If a person responds "No" to this statement, then further questions are not warranted at the time because the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) (American Psychiatric Association and others 2013) states that symptoms following a traumatic event should be present for at least one month to meet the definition of PTSD. Re-screening at a later date may be appropriate. There is a similar qualifying item in the PC-PTSD-5 that does not receive a score.

Table 176: Trauma Screen test items and items from the PC-PTSD-5. The following five items are from the PC-PTSD-5: nightmares_R, avoidsits_R, detached_R, selfblame_R, and watchful_R.

Item	Short Description (Response Categories)
traumevent_R	An extremely upsetting event occurred in my life at least one (1) month ago. Examples include serious accidents, witnessing or being the victim of violence, experiencing a natural disaster or war, or losing a loved one by traumatic means. If you are not sure whether the event was "extremely upsetting," respond yes to this question. *(No (if No, stop here); Yes (if Yes, continue); (no points value))*
avoidremind_R	During the past 7 days, I have tried to avoid people, places, or things that remind me of the traumatic event. *(No, disagree; Yes, agree; test item)*
avoidsits_R	In the past month, I have tried hard not to think about the event(s) or went out of my way to avoid situations that reminded me of the event(s). *(No, disagree=0; Yes, agree=1)*
avoidthoughts_R	During the past 7 days, I have tried to avoid thoughts or feelings about the traumatic event. *(No, disagree; Yes, agree; test item)*
awareness_R	During the past 7 days, I have a heightened awareness of the potential dangers surrounding me. *(No, disagree; Yes, agree; test item)*
detached_R	In the past month, I have felt numb or detached from people, activities, or my surroundings. *(No, disagree=0; Yes, agree=1)*
diffconc_R	During the past 7 days, I have found it difficult to concentrate.

Table 176: Trauma Screen test items and items from the PC-PTSD-5. The following five items are from the PC-PTSD-5: nightmares_R, avoidsits_R, detached_R, selfblame_R, and watchful_R. *(continued)*

Item	Short Description (Response Categories)
	(No, disagree; Yes, agree; test item)
disdreams_R	During the past 7 days, I have had distressing dreams that included elements of the traumatic event.
	(No, disagree; Yes, agree; test item)
distsleep_R	During the past 7 days, my typical sleep patterns have been disturbed since the traumatic event (such as restless sleep, difficulty falling asleep or staying asleep).
	(No, disagree; Yes, agree; test item)
flashback_R	During the past 7 days, I have had flashbacks that made me feel as if the traumatic event was occurring again.
	(No, disagree; Yes, agree; test item)
impfunc_R	During the past 7 days, I have experienced impaired functioning in either my work or in my relationships (or both).
	(No, disagree; Yes, agree; test item)
intrusive_R	During the past 7 days, distressing memories of the traumatic event have intruded on my thoughts.
	(No, disagree; Yes, agree; test item)
lessinterest_R	During the past 7 days, I have much less interest in participating in significant activities since the traumatic event.
	(No, disagree; Yes, agree; test item)
moredist_R	During the past 7 days, I have felt more distant and estranged from others since the traumatic event.
	(No, disagree; Yes, agree; test item)
negbeliefs_R	During the past 7 days, I have had persistent negative beliefs and expectations about my life that are more extreme than those I had before the traumatic event.
	(No, disagree; Yes, agree; test item)
negfeels_R	During the past 7 days, I have had strong negative feelings about myself such as guilt, fear, horror, anger, shame, etc.
	(No, disagree; Yes, agree; test item)
nightmares_R	In the past month, I have had had nightmares about the event(s) or thought about the event(s) when I did not want to.
	(No, disagree=0; Yes, agree=1)
noposemot_R	During the past 7 days, I have noticed an inability to feel positive emotions (such as happiness, satisfaction, hope, joy, etc.)
	(No, disagree; Yes, agree; test item)
outburst_R	During the past 7 days, I have had at least one episode of an angry outburst with little to no cause.
	(No, disagree; Yes, agree; test item)
physreact_R	During the past 7 days, things that reminded me of the traumatic event caused me to react physically (such as sweating or racing heart).
	(No, disagree; Yes, agree; test item)

Table 176: Trauma Screen test items and items from the PC-PTSD-5. The following five items are from the PC-PTSD-5: nightmares_R, avoidsits_R, detached_R, selfblame_R, and watchful_R. *(continued)*

Item	Short Description (Response Categories)
reminded_R	During the past 7 days, things that reminded me of the traumatic event caused me distress, either intensely or for a prolonged period. *(No, disagree; Yes, agree; test item)*
selfblame_R	In the past month, I have felt guilty or unable to stop blaming myself or others for the event(s) or any problems the event(s) may have caused. *(No, disagree=0; Yes, agree=1)*
startled_R	During the past 7 days, I have reacted strongly when something startled me. *(No, disagree; Yes, agree; test item)*
watchful_R	In the past month, I have been constantly on guard, watchful, or easily startled. *(No, disagree=0; Yes, agree=1)*

Instructions The instructions in this self-report screening are longer than for the other COMPAS-R scales. They are intended to acknowledge the emotions of the person responding and provide assurance that they need not rush through the assessment.

For the qualifying question, the instructions are:

- *Please read the following statement and decide whether you agree with it.*

The instructions for the following 23 items are:

- *Each of the following statements refers to the traumatic event(s) but pertains only to your experiences during the time stated (for example, the past 7 days).*
- *Please read each statement carefully and decide whether you agree or disagree. Take your time; there are no 'right or wrong' answers.*
- *Some of these items may give rise to strong emotional responses.*
- *Although every statement requires a response, feel free to skip over statements and return to them later.*
- *If you would like clarification of any of the statements, please ask the person administering this assessment. They are trained to answer questions about this screening tool.*
- *If you are unable to finish the questionnaire for any reason, you can arrange a time to continue later.*

Item inclusion Table 177 organizes the COMPAS-R Trauma Screen test items and the PC-PTSD-5 items by the domains of PTSD specified in the DSM-5, using their A-G labeling. Other PTSD screening tools were also examined. These were the SPAN: Self-Report Screen (4 items) (Davidson 2002), the Short Post-Traumatic Stress Disorder Rating Interview (SPRINT, 8 items) (Connor and Davidson 2001), and the Trauma Screening Questionnaire (TSQ, 10 items) (Brewin et al. 2002). The distribution and content of the COMPAS-R Trauma Screen test items reflects the prevalence of each domain from these screening tools.

Table 177: Trauma Screen test items and items from the PC-PTSD-5 organized by PTSD domain in DSM-5.

	DSM-5 Domain	PC-PTSD-5 items	COMPAS-R Trauma Screen test items
A/F	Stressor/Duration	item similar to traumevent	traumevent
B	Intrusion Symptoms	nightmares	intrusive disdreams flashback reminded physreact
C	Avoidance	avoidsits	avoidthoughts avoidremind
D	Negative mood and cognitions alterations	detached selfblame	negbeliefs negfeels lessinterest moredist noposemot
E	Alterations in arousal and reactivity	watchful	outburst awareness diffconc distsleep startled
G	Functioning		impfunc

Data collection and analysis Sufficient data will have been collected when we have 200 sets of responses from both Men and Women from at least three agencies. That is, 400 observations, evenly balanced between Men and Women respondents, from three agencies each, for a total of 1,200 samples. Each observation will consist of responses to all 24 items.

To obtain coverage of PTSD domains B, C, D, E, and G (see Table 177), correlations of the total PC-PTSD-5 score with scores obtained by combinations of test items representing the 5 domains will be examined. We will select the combination that optimizes the correlation. Further analysis will be performed to determine an appropriate cut point for the likelihood of PTSD.

It is hoped that the resulting screen will have some advantages for agencies that use it. First, it will have been developed using population-relevant data. Second, the wording of the items following the qualifying item emphasizes the past seven days, to obtain a greater degree of timeliness. Last, the resulting screen will include the respondent's own reflection of their day-to-day functioning.

8 Supervision Levels

COMPAS-R Core Supervision Levels are determined by the risk scores from the Summative GRRS, without consideration of whether the offense or arrest was opioid related, unless the Probation or Composite norm is used. Then, agencies can configure the scale set to move a person with an affirmative response to this item to the highest supervision level; this is done to help ensure a person's success in their assigned drug treatment. We describe three schema for determining supervision levels. There is no conviction version of the Supervision Levels at this time because there is no conviction version of the Summative GRRS. No one in the risk development sample scored below 11 or above 56, so the risk probabilities from the resulting model are not shown. 96% of the people in that sample scored between 20 and 48 on the Summative GRRS, so risk probability estimates outside of that range should be viewed with caution.

This section contains three figures (Figures 31, 33, and 34), each of which contains 5 columns.

- Column 1 shows the values of the Summative GRRS scores (risk contributions) in the development sample; this column is identical in all three figures.
- Column 2 shows the Recidivism probability estimate from the risk model; this column is also identical in all three figures.
- Column 3 shows the supervision levels for Men in the relevant scheme. These columns also indicate the level-wide recidivism rate of Men Probationers from the test subset of the risk data set. There are 41,769 probationers in this set, of which 32,260 are men.
- Column 4 shows the supervision levels Women in the relevant scheme. These columns also indicate the level-wide recidivism rate of Women Probationers from the test subset of the risk data set. 9,509 of the Probationers in this set are Women.
- Column 5 contains a plot showing the distribution of Probationers' scores from the test subset of the risk data set; this column is identical in all three figures.

8.1 Three Level Scheme

The Three Level Scheme indicates the cut points appropriate for an agency that has three supervision levels. The selection ratios for the three level scheme are 40% in Low, 30% in Medium, and 30% in High, for both Men and Women. Table 178 shows the score ranges from the Summative GRRS that determine the levels for the different norm sets.

Table 178: Summative GRRS score ranges for the COMPAS-R Three Level Supervision scheme.

	Level		
	Low	Medium	High
Male Composite	0 – 35	36 – 40	41 – 71
Male Prison	0 – 35	36 – 40	41 – 71
Male Probation	0 – 34	35 – 40	41 – 71
Male Jail (Client-specific)	0 – 38	39 – 42	43 – 71
Female Composite	0 – 33	34 – 40	41 – 71
Female Prison	0 – 35	36 – 40	41 – 71
Female Probation	0 – 31	32 – 38	39 – 71
Female Jail (Client-specific)	0 – 38	39 – 43	44 – 71

Figure 31 shows the estimated recidivism rates of Probationers for specific scores based on the risk model, with the actual recidivism rate from the Probation sample for each level (Low, Medium, High) by gender. These rates do not take into account whether the offense or arrest was opioid related. *When the Probation or Composite norm is selected during the scale set configuration process, agencies can opt for increasing the supervision level to High when the current offense is opioid related, in order to facilitate the person's chance of success in their assigned substance use treatment program.*

If a male probationer scored a 21 on the Summative GRRS, the risk model suggests he has a 14% chance of recidivating. That is, the model says that, on average, 14% of Probationers who take the Summative GRRS and obtain a score of 21 will have a new arrest within three years of community start. However, as the saying by the statistician George Box goes, "Models, of course, are never true, but fortunately it is only necessary that they be useful," (Box 1979). It may also be useful to an agency to know what the empirical, level-wide recidivism rate from the Probation sample is, in addition to the rate suggested by the model for a single score. We provide both in this document, although the software reports only the level-wide recidivism rate along with the supervision level. In the case of the man in our example, he scored a 21 which is among the lowest 40% of probationers' scores. 28% of the male probationers in Low recidivated within three years. In the software, his supervision level will be reported as Low if the agency is using the Three Level Scheme, along with a recidivism rate of 28%.

Risk contribution	Recidivism Probability Estimate (from risk model)	3 Supervision Levels	
		Men	Women
11 – 12	6%		
13	7%		
14 – 15	8%		
16	9%		
17	10%		
18	11%		
19	12%		
20	13%		
21	14%		
22	16%		
23	17%		
24	18%		
25	20%		
26	22%	40% in LOW 28% new arrest in 3 years	40% in LOW 24% new arrest in 3 years
27	23%		
28	25%		
29	27%		
30	29%		
31	31%		
32	34%		
33	36%		
34	38%		
35	41%	30% in MEDIUM 46% new arrest in 3 years	30% in MEDIUM 42% new arrest in 3 years
36	43%		
37	45%		
38	48%		
39	50%		
40	53%		
41	55%	30% in HIGH 62% new arrest in 3 years	30% in HIGH 58% new arrest in 3 years
42	58%		
43	60%		
44	63%		
45	65%		
46	67%		
47	69%		
48	71%		
49	73%		
50	75%		
51	77%		
52	79%		
53	80%		
54	82%		
55	83%		
56	85%		

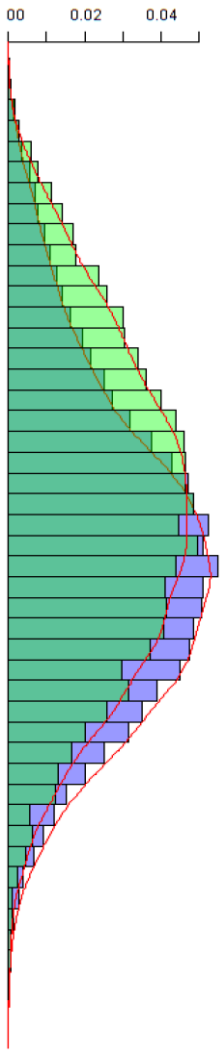


Figure 31: Estimated recidivism rates of Probationers for specific scores based on the risk model. The level-wide recidivism rate for each supervision level (Low, Medium, High) by gender is shown within the corresponding blue or green cells. The estimated recidivism rates do not take into account whether the offense or arrest was opioid related. *If the Probation or Composite norm is used, agencies can configure the scale set to move a person with a current offense that is opioid related to the highest supervision level in order to ensure the person's success in their assigned drug treatment.* In the histogram/density plot, Men's scores from the Probation test sample are represented by the blue bars, and Women's scores are represented by the green bars.

8.2 Four Level Scheme

The Four Level Scheme indicates the cut points appropriate for an agency that has four supervision levels. The selection ratios for the Four Level Scheme are 38% in Low, 21% in Medium, 19% in Medium-High, and 22% in High, for Men, and 44% in Low, 22% in Medium, 19% in Medium-High, and 15% in High for Women. Table 179 shows the score ranges from the Summative GRRS that determine the levels for the different norm sets. The selection ratios are most closely aligned with the selection ratios observed using the COMPAS Supervision Matrix from the standard COMPAS, shown in Figure 32. The COMPAS-R does not have a risk scale to measure a person's potential for violent recidivism. However, a person's scores the History of Violence and History of Non-compliance scales may be taken into consideration in conjunction with their Summative GRRS to help practitioners make informed decisions about supervision levels, (see section 4.1.5 of the Practitioner's Guide to COMPAS Core, (Northpointe Inc. 2019), *On Counter-Intuitive Predictions*).

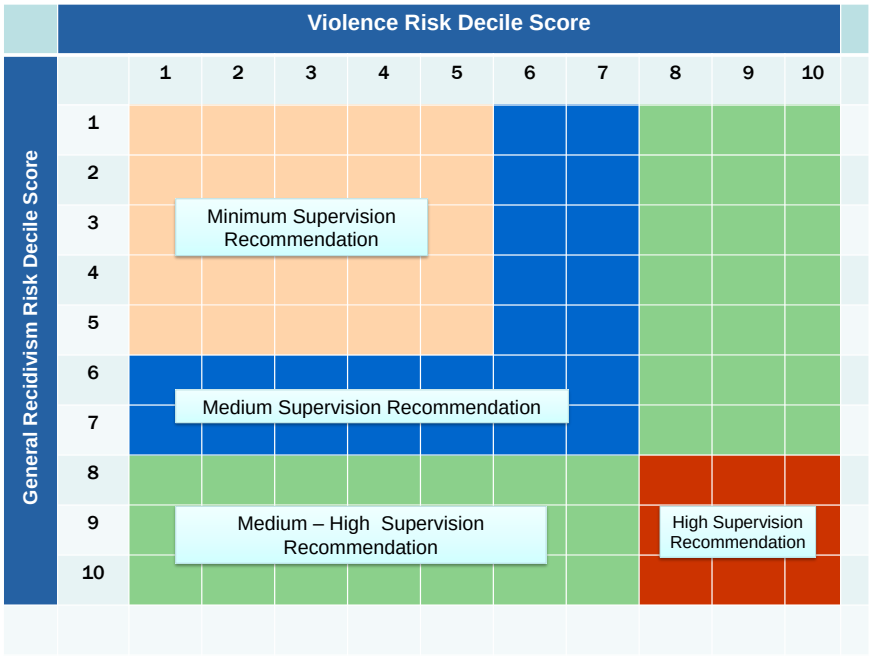


Figure 32: The Standard COMPAS Supervision Matrix, shown, takes into account a person's scores on both the General and Violent Recidivism Risk Scales. This matrix is not used with the COMPAS-R, although the selection ratios od the Four Level Scheme are closely aligned with the selection ratios observed using the COMPAS Supervision Matrix.

Continuing the example from section 8.1 with the male probationer who scored a 21 on the Summative GRRS, the risk model still suggests he has a 14% chance of recidivating because it is the same model. His score is among the lowest 38% of probationers' scores. 28% of the male probationers in Low will recidivate within three years. In the software, his supervision level will be reported as Low if the agency is using the Four Level Scheme, along with a recidivism rate of 28%. If this person's History of Non-compliance or History of Violence score is High, it may be advisable for the practitioner to override the scheme and recommend a supervision level higher than Low.

Table 179: Summative GRRS score ranges for the COMPAS-R Four Level Supervision scheme.

	Level			
	Low	Medium	Medium-High (with override to High consideration)	High
Male Composite	0 – 35	36 – 39	40 – 42	43 – 71
Male Prison	0 – 35	36 – 38	39 – 42	43 – 71
Male Probation	0 – 34	35 – 38	39 – 41	42 – 71
Male Jail (Client-specific)	0 – 38	39 – 40	41 – 43	44 – 71
Female Composite	0 – 34	35 – 39	40 – 43	44 – 71
Female Prison	0 – 36	37 – 40	41 – 43	44 – 71
Female Probation	0 – 32	33 – 37	38 – 42	43 – 71
Female Jail (Client-specific)	0 – 39	40 – 42	43 – 45	46 – 71

Figure 33 shows the estimated recidivism rates of Probationers for specific scores based on the risk model, with the actual recidivism rate from the Probation sample for each level (Low, Medium, Medium-High, High) by gender. These rates do not take into account whether the offense or arrest was opioid related. *When the Probation or Composite norm is selected during the scale set configuration process, agencies can opt for increasing the supervision level to High when the current offense is opioid related, in order to facilitate the person's chance of success in their assigned substance use treatment program.*

Risk contribution	Recidivism Probability Estimate (from model)	4 Supervision Levels	
		Men	Women
11 – 12	6%		
13	7%		
14 – 15	8%		
16	9%		
17	10%		
18	11%		
19	12%		
20	13%		
21	14%		
22	16%		
23	17%		
24	18%		
25	20%		
26	22%		
27	23%	38% in LOW	44% in LOW
28	25%		
29	27%	28% new arrest in 3 years	25% new arrest in 3 years
30	29%		
31	31%		
32	34%		
33	36%		
34	38%		
35	41%	21% in MEDIUM	22% in MEDIUM
36	43%		
37	45%	44% new arrest in 3 years	43% new arrest in 3 years
38	48%		
39	50%	19% in MEDIUM-HIGH	19% in MEDIUM-HIGH
40	53%		
41	55%	53% new arrest in 3 years	52% new arrest in 3 years
42	58%		
43	60%		
44	63%		
45	65%		
46	67%	22% in HIGH	15% in HIGH
47	69%		
48	71%	64% new arrest in 3 years	65% new arrest in 3 years
49	73%		
50	75%		
51	77%		
52	79%		
53	80%		
54	82%		
55	83%		
56	85%		

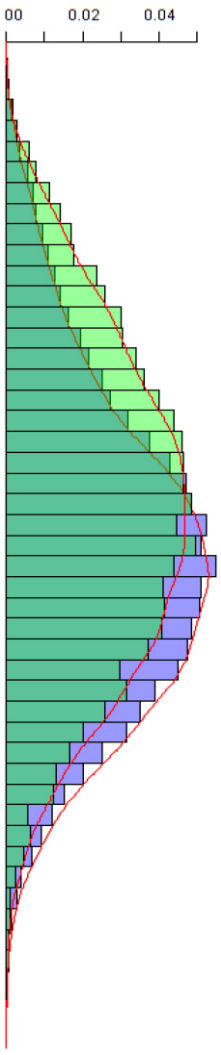


Figure 33: Estimated recidivism rates of Probationers for specific scores based on the risk model. The level-wide recidivism rate for each supervision level (Low, Medium, Medium-High, High) by gender is shown within the corresponding blue or green cells. The estimated recidivism rates do not take into account whether the offense or arrest was opioid related. *If the Probation or Composite norm is used, agencies can configure the scale set to move a person with a current offense that is opioid related to the highest supervision level in order to ensure the person's success in their assigned drug treatment.* In the histogram/density plot, Men's scores from the Probation test sample are represented by the blue bars, and Women's scores are represented by the green bars.

8.3 Five Level Scheme

The Five Level Scheme was designed to comply as closely as possible with the guidance provided in *A Five-Level Risk and Needs System: Maximizing Assessment Results in Corrections Through the Development of a Common Language*, (Hanson et al. 2017). Some features of our model made an exact implementation impossible. First, the authors recommend that people in the lowest level have an expected recidivism rate of 5% or lower, and those in the highest level have an expected recidivism rate of at least 85%. Figure 34 makes it clear that our data did not support building a scheme that matched these particular requirements because no one in the development sample had scores at either extreme of the model. Another difference is that the COMPAS-R definition of recidivism covers a three year follow time, while theirs covers only a two year period. Despite these differences, score ranges and recidivism probabilities for levels II–IV, match the requirements of the document more robustly: approximately 30% of those in Level II, about 45% in Level III, and about 60% in Level IV recidivate within three years. Hanson et al. suggest a two year recidivism rate of 5%–29% for those in Level II, 30%–49% for those in Level III, and 50%–84% for those in Level IV.

Table 180: Summative GRRS score ranges for the COMPAS-R Five Level Supervision scheme.

	Level				
	I	II	III	IV	V
Male Composite	0 – 24	25 – 34	35 – 42	43 – 46	47 – 71
Male Prison	0 – 23	24 – 33	34 – 42	43 – 46	47 – 71
Male Probation	0 – 23	24 – 33	34 – 42	43 – 46	47 – 71
Male Jail (Client-specific)	0 – 31	32 – 37	38 – 43	44 – 47	48 – 71
Female Composite	0 – 22	23 – 33	34 – 41	42 – 46	47 – 71
Female Prison	0 – 22	23 – 34	35 – 42	43 – 46	47 – 71
Female Probation	0 – 21	22 – 31	32 – 40	41 – 45	46 – 71
Female Jail (Client-specific)	0 – 29	30 – 38	39 – 44	45 – 47	48 – 71

Figure 34 shows the estimated recidivism rates of Probationers for specific scores based on the risk model, with the actual recidivism rate from the Probation sample for each level (I, II, III, IV, V) by gender. These rates do not take into account whether the offense or arrest was opioid related. *When the Probation or Composite norm is selected during the scale set configuration process, agencies can opt for increasing the supervision level to the highest level, V, when the current offense is opioid related; this is done to help facilitate the person's chance of success in their assigned substance use treatment program.*

Risk contribution	Recidivism Probability Estimate (from model)	5 Supervision Levels	
		Men	Women
11 – 12	6%	6% in Level I 14% new arrest in 3 years	5% in Level I 14% new arrest in 3 years
13	7%		
14 – 15	8%		
16	9%		
17	10%		
18	11%		
19	12%		
20	13%		
21	14%		
22	16%		
23	17%		
24	18%		
25	20%	27% in Level II 30% new arrest in 3 years	32% in Level II 26% new arrest in 3 years
26	22%		
27	23%		
28	25%		
29	27%		
30	29%		
31	31%		
32	34%		
33	36%		
34	38%	46% in Level III 47% new arrest in 3 years	42% in Level III 44% new arrest in 3 years
35	41%		
36	43%		
37	45%		
38	48%		
39	50%		
40	53%		
41	55%		
42	58%		
43	60%	16% in Level IV 63% new arrest in 3 years	16% in Level IV 58% new arrest in 3 years
44	63%		
45	65%		
46	67%		
47	69%	5% in Level V 72% new arrest in 3 years	5% in Level V 69% new arrest in 3 years
48	71%		
49	73%		
50	75%		
51	77%		
52	79%		
53	80%		
54	82%		
55	83%		
56	85%		

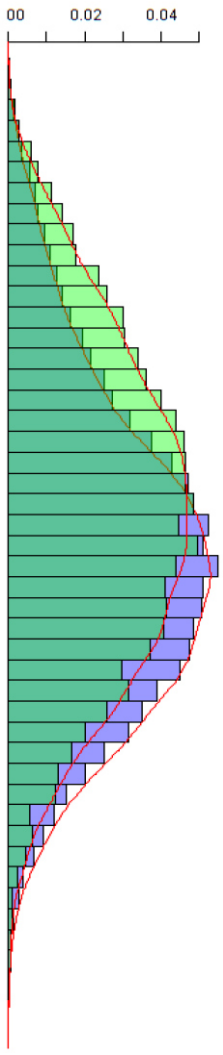


Figure 34: Estimated recidivism rates of Probationers for specific scores based on the risk model. The level-wide recidivism rate for each supervision level (I, II, III, IV, V) by gender is shown within the corresponding blue or green cells. The estimated recidivism rates do not take into account whether the offense or arrest was opioid related. *If the Probation or Composite norm is used, agencies can configure the scale set to move a person with a current offense that is opioid related to the highest supervision level (V) in order to ensure the person's success in their assigned drug treatment.* In the histogram/density plot, Men's scores from the Probation test sample are represented by the blue bars, and Women's scores are represented by the green bars.

9 Tests of Validity

9.1 Random Response Test

The Random Response Test is designed to identify respondents who may be responding randomly or in a highly inconsistent or careless manner. The test is modified from the test in the standard COMPAS Core; that test was designed to be similar to the VRIN test that appears in the MMPI-2 (Graham 2000).

Item Pairs The random response scale is based on 16 highly correlated pairs of COMPAS-R Core scale items. These pairs appear in Table 181. All of these items are described in tables in the previous sections of this report. Note that some items appear repeatedly among the pairs. For example, the item *local_gangs* is the second item in both the *citizens_weapons/local_gangs* pair and the *local_crime/local_gangs* pair.

Some items in the table end with "b1." They are binary items derived from the items with corresponding names that are recoded to binary with the following rules: responses of 0 and 1 are recoded to 0 and responses of 2 and 3 are recoded to 1. These items are from the *Thinking and Attitudes*, and *Personality Traits/Anger* scales.

Some items in the table end with "b2." They are binary items derived from the items with corresponding names that are recoded to binary with the following rules: responses of 0 remain 0 and responses of 1 and 2 are recoded to 1. These items are from the *Substance Use* and *Socialization History* scales.

Calculating Scores The steps involved in calculating scores are as follows:

1. Calculate the binary versions of the items that are listed in Table 181
2. Count the number of values that mismatch for the 16 pairs. For example, a mismatch occurs if a respondent gives a 0 for the first item in a pair and a 1 for the second item.
3. Divide the number of mismatches by 16.

Cutting Score Placement The cutting score for the Random Responding scale is placed at 0.4375. This allows one to flag respondents who score in the upper 5% of the distribution of mismatches from our norms data. This is equivalent to having 7 or more mismatches among the 16 pairs.

Table 181: Random Response Test Item Pairs.

Item 1	Item 2
haveempschool	job
parent_jailed	conv_parent - b2
trouble_no - b1	law_nochance - b1
citizens_weapons	local_crime
citizens_weapons	local_gangs
local_crime	local_gangs
insults - b1	threaten - b1
never_temper - b1	short_temper - b1
threaten - b1	angry_dangerous - b1
violent_person - b1	short_temper - b1
angry_dangerous - b1	short_temper - b1
angry_dangerous - b1	violent_person - b1
ad_arrest - b2	benefit_rx_ad - b2
ad_arrest - b2	ever_rx_ad - b2
ever_rx_ad - b2	benefit_rx_ad - b2
ad_arrest - b2	blame_drugs

9.2 Internal validation checks

Internal validation checks refers to item response logic. Internal validation checks are used to confirm that appropriate inputs are used when the input requires the user to type text or insert a number; they also confirm that response options make sense when compared to other response options. Items responses can contradict other item responses for several reasons:

- Questions are misread or misunderstood
- Responses are random (discussed above)
- The incorrect response was selected
  - the wrong radio button was selected
  - the wrong symbol/key is used on items that require users to manually input responses

The software contains many different internal validation checks. For example, one question on the COMPAS-R Core assessment requires users to insert their client's age when they were first arrested. The input required for this field is a text box, where the user must enter a number; this number must be greater than ten. If it is not greater than ten, a warning will appear before the assessment can be computed alerting the user that for most calculations including age, the age entered will be rounded to ten. If the user fails to even input a number, another warning will alert the user that an integer must be entered for the age at first field. Both *warnings* or *alerts*, are types of internal validation checks.

Another type of validation check confirms that related items make sense. For example the two items from the Residential Instability scale in COMPAS-R Core, ***res_moves*** and ***yrs_address***, in Table 182 could have some some contradicting item responses. For instance, if the first response (None) is selected for the item ***res_moves***, but the first response (0 - 5 months) is also selected for the second item, a warning will alert the user that their client could not have actually only lived at their current address for less than 6 months if they have also not moved within the last year.

Table 182: Two items from the Residential Instability scale with item descriptions and response options.

	Description	Response Category
res_moves_R	How often have you moved in the last 12 months?	None|1 time|2 times|3 times|4 times|5 or more times
yrs_address_R	How long have you lived at your current address?	0-5 months|6-11 months|1-3 years|4-5 years|6 years or longer

These type of checks on item responses are referred to as *internal validation checks*. There are many more throughout the software for both the standard and conviction versions of the COMPAS-R Core applications.

10 Appendix

10.1 Normative and Test Samples

The COMPAS Core data were provided by four state Departments of Corrections (DOCs) and a single large Jail. The DOCs provided data from Prison (Incarcerated), Parole, and Probation populations.

In the COMPAS Core sample, a large proportion of the assessments (80.04%) were from the Probation population. Assessments drawn from the Prison (Incarcerated) population represent 8.01% of the sample, 5.17% of the sample are from the Parole population, 0.46% of the sample are from the Jail population, and 6.32% of the sample belong to populations that were difficult to identify specifically. Their presence in the sample allows some noise to enter the model development that can enhance the generalizability of the scales.

Table 183: Prison and Parole are combined to form the Prison norm

	Normative Sample			Test Sample	
	Arrest	Conviction	All	Arrest	Conviction
Jail	2,306		2,306		
Parole	2,724		2,724	23,130	
Prison	2,346	2,346	4,692	35,371	
Probation	8,440	2,963	11,403	287,876	101,089
Composite	15,816		21,125		

Additionally, the Prison (Incarcerated), Parole, and Probation subgroups of the sample can be further partitioned by whether individuals were assessed using the "standard version" or "conviction version" of a scale set. These scale sets are distinguished from each other by whether they contain the standard version or conviction version of the following scales: GRRS, VRRS, Cognitive Behavioral, Criminal Involvement, History of Non-Compliance, History of Violence, Social Adjustment, and Socialization Failure. None of the other scales rely on either arrest or conviction data; that is, the remaining scales are the same regardless of whether they belong to a scale set designated as "standard version" or "conviction version."

Tables 185 and 186 show the numbers and overall percentages of the sample by the populations from which the data were drawn and the recidivism risk assessments they received. "Other" represents individuals for whom no recidivism risk assessment score was available. However, these people received need assessments. No one from the Jail sample received a conviction version of any scales.

Table 184: Distribution of the populations in the Composite Norm group

Population	Frequency	Percent
Jail	2306	11
Prison and Parole	7416	35
Probation	11403	54
Total	21125	100

Given the large size of the sampling frames and the need to have a development, or normative, sample that adequately represents the different populations, we obtained a development sample of size 21,123, composed of: all 2,306 individuals from the Jail sampling frame who received the GRRS (or Other); 2,724 randomly selected individuals from the Parole sampling frame who received standard versions of the scales; 2,346 randomly selected individuals from the Prison (Incarcerated) sampling frame who received standard versions of the scales and all 2,346 randomly selected individuals from the Prison (Incarcerated) sampling frame who received conviction versions of the scales; 8,440 randomly selected individuals from the Probation sampling frame who received standard versions of the COMPAS scales; and 2,963 randomly selected individuals from the Probation sampling frame who received conviction versions of the scales. This yields a development sample with 11% from Jail, 35% from Prison (combined Prison (Incarcerated) and Parole), and 54% from Probation, a composition similar to that found in the corrections population nationwide (Kaeble and Cowhig 2018). See Table 184 for a summary of populations in the Composite Norm group.

The development sample will be examined to ensure it is demographically similar to the remaining test sample. The test sample consists of both arrest and conviction data for Probationers, arrest data only for Prison (Incarcerated), and no data for Jail inmates or Parolees.

The next several sections will show the demographic features of the selected samples for the different populations (Jail, Prison (Incarcerated), Parole, Probation), followed by demographic features of the composite sample for non-conviction scales.

Table 185: Counts of unique individuals whose earliest assessments comprise the COMPAS-R sample: 0.46% of the data are from Jail, 5.17% of the data are from Parole, 8.01% of the data are from Prison, 80.04% of the data are from Probation, and 6.32% of the data are from unidentified populations.

	Population									
	Jail		Parole		Prison		Probation		Unidentified	
	arrest	conviction	arrest	conviction	arrest	conviction	arrest	conviction	arrest	conviction
GRRS (arrest)	2302	0	25854	0	37714	0	296253	0	31605	0
GRRS (conviction)	0	0	0	0	0	1513	0	40946	0	0
Other	4	0	0	0	3	833	63	63108	1	0

Table 186: Overall percentages of unique individuals whose earliest assessments comprise the COMPAS-R sample.

	Population									
	Jail		Parole		Prison		Probation		Unidentified	
	arrest	conviction	arrest	conviction	arrest	conviction	arrest	conviction	arrest	conviction
GRRS (arrest)	0.46	0	5.17	0	7.54	0.00	59.23	0.00	6.32	0
GRRS (conviction)	0.00	0	0.00	0	0.00	0.30	0.00	8.19	0.00	0
Other	0.00	0	0.00	0	0.00	0.17	0.01	12.62	0.00	0

10.1.1 Jail Data

All Jail data is selected for inclusion in the development sample from the full sampling frame.

Demographic information for Jail inmates who took the complete COMPAS Core assessment (or Other) is displayed in Tables 187–188.

Table 187: Ethnicity percentages by gender for the development subsample of Jail inmates who took the complete COMPAS Core.

	Black	Hispanic	Other	White
Female	34.32	1.23	0.49	63.95
Male	52.21	1.42	0.47	45.89
Overall	49.05	1.39	0.48	49.09

Table 188: Gender percentages by ethnicity for the development subsample of Jail inmates who took the complete COMPAS Core.

	Black	Hispanic	Other	White	Overall
Female	12.29	15.62	18.18	22.88	17.56
Male	87.71	84.38	81.82	77.03	82.39

10.1.2 Prison (Incarcerated) Data

Demographic information for the development samples of incarcerated persons who took the complete COMPAS Core assessment (standard version) is displayed in Tables 189–190. Demographic information for the development samples of incarcerated persons who took the complete COMPAS Core assessment (conviction version) is displayed in Tables 191–192. Demographic information for the development samples of incarcerated persons who took the complete COMPAS Core assessment (either version) is displayed in Tables 193–194. Demographic information for the **test** sample of incarcerated persons who took the complete COMPAS Core assessment (or Other, standard version only) is displayed in Tables 195–196.

Development data: standard version

Table 189: Ethnicity percentages by gender for the randomly selected development subsample of incarcerated persons who took the complete COMPAS Core (standard version).

	Black	Hispanic	Other	White
Female	30.00	30.00	10.00	30.00
Male	30.01	30.01	9.97	30.01
Overall	30.01	30.01	9.97	30.01

Table 190: Gender percentages by ethnicity for the randomly selected development subsample of incarcerated persons who took the complete COMPAS Core (standard version).

	Black	Hispanic	Other	White	Overall
Female	20.03	20.03	20.09	20.03	20.03
Male	79.97	79.97	79.91	79.97	79.97

Development data: conviction version

Table 191: Ethnicity percentages by gender for the development subsample of incarcerated persons who took the complete COMPAS Core (conviction version).

	Black	Hispanic	Other	White
Female	41.30	1.09	2.90	54.71
Male	50.77	2.61	2.32	44.30
Overall	49.66	2.43	2.39	45.52

Table 192: Gender percentages by ethnicity for all the incarcerated persons who took the complete COMPAS Core (conviction version).

	Black	Hispanic	Other	White	Overall
Female	9.79	5.26	14.29	14.14	11.76
Male	90.21	94.74	85.71	85.86	88.24

Development data: arrest and conviction versions combined

Table 193: Ethnicity percentages by gender for all the incarcerated persons who took the complete COMPAS Core (combined).

	Black	Hispanic	Other	White
Female	34.18	19.30	7.37	39.14
Male	40.90	15.64	5.96	37.51
Overall	39.83	16.22	6.18	37.77

Table 194: Gender percentages by ethnicity for the development subsample of incarcerated persons who took the complete COMPAS Core (combined).

	Black	Hispanic	Other	White	Overall
Female	13.64	18.92	18.97	16.48	15.9
Male	86.36	81.08	81.03	83.52	84.1

Test data: standard version only

Stratified sampling was used to guarantee that 20% of the sample of incarcerated persons who took the standard version of the COMPAS scales were women (see Table 190). Since the overall proportion of female incarcerated persons in the data is only 9.35% , the test data contains a smaller proportion of women (see Table 196).

Table 195: Ethnicity percentages by gender for the test subsample of incarcerated persons who took the complete COMPAS Core (standard version).

	Black	Hispanic	Other	White
Female	25.03	10.10	3.93	60.93
Male	41.57	14.19	3.65	40.59
Overall	40.17	13.84	3.67	42.32

Table 196: Gender percentages by ethnicity for the test subsample of incarcerated persons who took the complete COMPAS Core (standard version).

	Black	Hispanic	Other	White	Overall
Female	5.29	6.19	9.08	12.21	8.48
Male	94.71	93.81	90.92	87.79	91.52

10.1.3 Parole Data

Demographic information for the development samples of parolees who took the complete COMPAS Core assessment (standard version) is displayed in Tables 197–198. Demographic information for the **test** sample of parolees who took the complete COMPAS Core assessment (or Other, standard version only) is displayed in Tables 199–200. There were very few people in the Parole data that were identified as Hispanic, so the development sample was chosen to strike a balance between Blacks and Whites.

Development data: standard version

Table 197: Ethnicity percentages by gender for the randomly selected development subsample of parolees who took the complete COMPAS Core (standard version).

	Black	Hispanic	Other	White
Female	47.71	0.37	4.22	47.71
Male	46.31	2.29	5.09	46.31
Overall	46.59	1.91	4.92	46.59

Table 198: Gender percentages by ethnicity for the randomly selected development subsample of parolees who took the complete COMPAS Core (standard version).

	Black	Hispanic	Other	White	Overall
Female	20.49	3.85	17.16	20.49	20.01
Male	79.51	96.15	82.84	79.51	79.99

Test data: standard version

Table 199: Ethnicity percentages by gender for the test subsample of parolees who took the complete COMPAS Core (standard version).

	Black	Hispanic	Other	White
Female	31.55	0.04	0.56	67.85
Male	56.65	0.00	0.24	43.11
Overall	53.55	0.01	0.28	46.16

Table 200: Gender percentages by ethnicity for the test subsample of parolees who took the complete COMPAS Core (standard version).

	Black	Hispanic	Other	White	Overall
Female	7.26	50	25	18.10	12.32
Male	92.74	50	75	81.89	87.68

10.1.4 Probation Data

Demographic information for the development samples of Probationers who took the complete COMPAS Core assessment (or Other) is displayed in Tables 201–201. Demographic information for the test sample of Probationers who took the complete COMPAS Core assessment (or Other, standard version only) is displayed in Tables 201–201.

Development data: standard version

Table 201: Ethnicity percentages by gender for the randomly selected development subsample of Probationers who took the complete COMPAS Core (standard version).

	Black	Hispanic	Other	White
Female	30	30	9.99	30
Male	30	30	10.00	30
Overall	30	30	10.00	30

Table 202: Gender percentages by ethnicity for the randomly selected development subsample of probationers who took the complete COMPAS Core (standard version).

	Black	Hispanic	Other	White	Overall
Female	25	25	25	25	25
Male	75	75	75	75	75

Development data: conviction version

Table 203: Ethnicity percentages by gender for the development subsample of probationers who took the complete COMPAS Core (conviction version).

	Black	Hispanic	Other	White
Female	30.01	30.01	9.96	30.01
Male	30.00	30.00	10.00	30.00
Overall	30.00	30.00	9.99	30.00

Table 204: Gender percentages by ethnicity for all the probationers who took the complete COMPAS Core (conviction version).

	Black	Hispanic	Other	White	Overall
Female	25.08	25.08	25	25.08	25.08
Male	74.92	74.92	75	74.92	74.92

Development data: arrest and conviction versions combined

Table 205: Ethnicity percentages by gender for all the probationers who took the complete COMPAS Core (combined).

	Black	Hispanic	Other	White
Female	30	30	9.99	30
Male	30	30	10.00	30
Overall	30	30	10.00	30

Table 206: Gender percentages by ethnicity for the development subsample of probationers who took the complete COMPAS Core (combined).

	Black	Hispanic	Other	White	Overall
Female	25.02	25.02	25	25.02	25.02
Male	74.98	74.98	75	74.98	74.98

Test data: standard version only

Table 207: Ethnicity percentages by gender for the test subsample of probationers who took the complete COMPAS Core (standard version).

	Black	Hispanic	Other	White
Female	23.62	5.24	4.17	66.97
Male	33.80	6.01	3.45	56.74
Overall	31.21	5.81	3.65	59.32

Table 208: Gender percentages by ethnicity for the test subsample of probationers who took the complete COMPAS Core (standard version).

	Black	Hispanic	Other	White	Overall
Female	19.17	22.81	28.95	28.60	25.33
Male	80.81	77.17	70.41	71.38	74.62

Test data: conviction version only

Table 209: Ethnicity percentages by gender for the test subsample of probationers who took the complete COMPAS Core (conviction version).

	Black	Hispanic	Other	White
Female	32.08	1.55	2.07	64.31
Male	38.33	2.49	2.41	56.77
Overall	36.42	2.20	2.31	59.07

Table 210: Gender percentages by ethnicity for the test subsample of probationers who took the complete COMPAS Core (conviction version).

	Black	Hispanic	Other	White	Overall
Female	26.84	21.43	27.36	33.18	30.48
Male	73.16	78.57	72.64	66.82	69.52

10.1.5 Composite Data

Development data: standard version

Table 211: Ethnicity percentages by gender for the randomly selected Composite development subsample those who took the complete COMPAS Core (standard version).

	Black	Hispanic	Other	White
Female	33.23	22.12	8.02	36.63
Male	36.33	20.67	7.65	35.35
Overall	35.63	20.99	7.73	35.64

Table 212: Gender percentages by ethnicity for the randomly selected development subsample those who took the complete COMPAS Core (standard version).

	Black	Hispanic	Other	White	Overall
Female	20.81	23.52	23.14	22.94	22.32
Male	79.19	76.48	76.86	77.04	77.67

Development data: combined

Table 213: Ethnicity percentages by gender for the randomly selected Composite development subsample those who took the complete COMPAS Core (all versions).

	Black	Hispanic	Other	White
Female	33.19	22.14	8.02	36.65
Male	37.29	19.66	7.30	35.75
Overall	36.40	20.19	7.46	35.95

Table 214: Gender percentages by ethnicity for the randomly selected development subsample those who took the complete COMPAS Core (all version).

	Black	Hispanic	Other	White	Overall
Female	19.64	23.61	23.17	21.95	21.53
Male	80.36	76.39	76.83	78.04	78.46

The race and gender balances of the data sets described in the appendix changed during scale development due to cases being removed for missing data or responses that were not included in the final scales. Table 215 shows the distribution of cases by race and sex within the composite norm set for the final scales in COMPAS-R Core.

Table 215: Percent of cases by race and sex within the composite norm set for the final scales in COMPAS-R Core.

COMPAS-R Core Scale	Black	Hispanic	Other	White	Total
Summative GRRS- Women	35	25	5	36	23
Summative GRRS- Men	39	21	5	35	77
Legal System Involvement- Women	35	25	5	36	23
Legal System Involvement- Men	39	21	5	35	77
History of Non-Compliance- Women	35	25	5	36	23
History of Non-Compliance- Men	39	21	5	35	77
History of Violence- Women	33	22	8	37	22
History of Violence- Men	36	21	8	35	78
Current Violence- Women	36	14	9	41	20
Current Violence- Men	41	13	7	38	80
Associates and Peers- Women	36	14	9	41	20
Associates and Peers- Men	41	13	7	39	80
Cognitive Behavioral- Women	30	14	10	46	21
Cognitive Behavioral- Men	37	16	8	39	79
Society and Routines- Women	32	17	9	41	19
Society and Routines- Men	39	16	8	38	81
Leisure and Recreation- Women	36	14	8	42	20
Leisure and Recreation- Men	41	13	7	39	80
Thinking and Attitudes- Women	34	14	9	43	21
Thinking and Attitudes- Men	41	13	7	39	79
Personality Traits- Women	37	14	8	42	21
Personality Traits- Men	43	13	7	38	79
Anger- Women	37	14	8	42	21
Anger- Men	43	13	7	38	79
Socialization History- Women	32	13	11	44	20
Socialization History- Men	37	15	8	39	80
Financial- Women	31	20	11	38	18
Financial- Men	38	17	8	38	82
Vocational/Educational- Women	33	22	8	37	22
Vocational/Educational- Men	37	20	7	36	78
Residential Stability- Women	36	14	9	42	20
Residential Stability- Men	41	13	7	39	80
Substance Use- Women	38	12	8	42	21
Substance Use- Men	42	12	7	39	79

10.2 Transformation of the refitted General Recidivism Risk Scale to a summative scale

This section is intended to demonstrate the method by which the summative scoring scheme was obtained. Part of our effort in revising the COMPAS Core was to make the transition from standard COMPAS Core scoring to COMPAS-R Core scoring as non-disruptive to agencies as possible.

We present a logistic regression model for the general recidivism risk scale, refitted to the current recidivism data that has been made available to us through the generosity of several agencies. Then we demonstrate the steps taken to generate an integer value for the risk score, given the contributions of the various scale inputs.

The general approach to transforming the logistic regression equation to a summative scale is to modify each term of the re-fitted equation for the GRRS so that outputs are non-negative integers that can be easily added.

Specifically:

- For each term in the logistic regression equation, input the range of possible values associated with that term. For example, values from 0 to 15, the range of scores obtainable on the Legal System Involvement scale, would be input to the term $0.593(logcrimv)$.
- Multiply each value returned by 10, then round each result to the nearest whole number. This step facilitates the generation of easily added values.
- Finally, if the smallest value in the result of the previous step is not 0, adjust all the values by the same amount so that the smallest value is 0.

The re-fitted logit equation for the GRRS raw score is

$$\begin{aligned} newGRRSraw = &- 0.683 \\ &+ 0.593(logcrimv) \\ &+ 0.174(voced6) \\ &+ 0.489(drugprob5) \\ &- 0.378(age.1) \\ &- 0.613(logage1) \\ &+ 0.393(logarate), \end{aligned} \tag{1}$$

where

$$\begin{aligned} logcrimv &= \ln(\text{criminv} + 1) \\ voced6 &= \text{voced}/6 \\ drugprob5 &= (\text{drugprob} + 5)/5 \\ age.1 &= \ln(\text{age} - 14) \\ logage1 &= \ln(\text{age.first}) \\ logarate &= \ln(100 * \text{arrest.r}). \end{aligned}$$

- "criminv" is the raw score from the COMPAS-R Core Legal System Involvement scale. It takes values from 0 to 15.
- "voced" is the raw score from the COMPAS-R Core Vocational/Educational scale. It takes values from 0 to 13.
- "age" is the person's age at assessment, calculated using the formula (screening date – birth date)/365.25, which is then rounded to the nearest whole number. In practical terms, a person's age is considered the age they are at their *nearest* birthday.
- "age.first" is the age at which the person was first arrested. This is an integer reported from criminal history records.
- "drugprob" is the score from the COMPAS-R Core Drug Problems scale. It takes values from 0 to 5.
- "arrest.r" is the arrest rate. It is calculated using the formula (number of previous arrests + 1)/("age" – "age.first" + 1).

Legal System Involvement

The first transformation is for the Legal System Involvement scale. Each of the possible scores from 0 to 15 is evaluated using the expression in the GRRS equation, multiplied by 10, then rounded to the nearest whole number. Tables 216 and 217 show the results of the transformation. The smallest contribution was 0, so no other adjustment was needed.

Table 216: The possible raw scores from the Legal System Involvement scale are listed along with their transformations using the formula $0.593 \times \ln(\text{criminv} + 1)$, multiplying the result by 10, and rounding to the nearest whole number. Since more than one raw score can be mapped to the same risk contribution, a simpler table of these values is shown in Table 217.

Raw Score	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Risk contribution	0	4	7	8	10	11	12	12	13	14	14	15	15	16	16	16

Table 217: Simplified version of Table 216 displaying the ranges of raw scores from the Legal System Involvement scale listed along with their contributions to the new raw risk score.

Raw Score	Risk contribution
0	0
1	4
2	7
3	8
4	10
5	11
6 – 7	12
8	13
9 – 10	14
11 – 12	15
13 – 15	16

Vocational/Educational

The new Vocational/Educational scale scores will go from 0 to 13. Tables 218 and 219 show the new Vocational Educational scale scores to be used in COMPAS-R, and the risk contributions given those scores.

Table 218: The raw scores from the Vocational/Educational scale are listed along with their transformations from using the formula 0.174 ×voced/6, multiplying the result by 10, and rounding to the nearest whole number. Since more than one raw score can be mapped to the same risk contribution, a more compact table of these values is shown in Table 219.

New Raw Score	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Risk contribution	0	0	1	1	1	1	2	2	2	3	3	3	3	4

Table 219: Simplified version of Table 218 displaying the ranges of raw scores from the Vocational/Educational scale listed along with their contributions to the new raw risk score.

New Raw Score	Risk contribution
0 – 1	0
2 – 5	1
6 – 8	2
9 – 12	3
13	4

Drug Problems

The Drug Problems scale is composed of three items from the Substance Use scale and two additional items. The total possible points on this scale range from 0 to 5. Each of the possible scores are evaluated using the expression in the GRRS equation, multiplied by 10, rounded to the nearest whole number, then 5 is subtracted from each score so that the lowest possible risk contribution is 0. Table 220 shows the Drug Problems scale scores and the risk contribution for each score.

Table 220: The possible raw scores from the Drug Problems scale are listed along with their transformations using the formula 0.489 ×(drugprob + 5)/5, multiplying the result by 10, rounding to the nearest whole number. For ease of use, the risk contributions are shifted so that the smallest risk contribution is 0. This is accomplished by adding 5 to each possible risk contribution. 5 can be added back for comparing the new raw scores to the original raw scores, if desired.

Raw Score	0	1	2	3	4	5
Risk contribution	0	1	2	3	4	5

Age at assessment

Integers representing the range of assessment ages are evaluated using the expression in the GRRS equation, multiplied by 10, rounded to the nearest whole number, and then 16 is added to each score so that the smallest contribution is 0. Recall that a person's age is considered the age they are at their nearest birthday. Table 221 shows the risk contributions by age for corresponding age ranges. We include risk contribution values for ages 16 and 17 because lower age limits for adult

assessments vary across agencies, but these ages are assigned the same risk contribution as 18 year olds.

Table 221: Age ranges and their contributions to the new raw risk score. Transformations are obtained by using the formula -0.378 $\times$ ln(age -14), multiplying the result by 10, and rounding to the nearest whole number. For ease of use, the risk contributions are shifted so that the smallest risk contribution is 0. This is accomplished by adding 16 to each possible risk contribution. This amount can be subtracted back when necessary for comparing the new raw scores to the original raw scores.

Age	Risk contribution
16 – 18	11
19	10
20 – 21	9
22 – 23	8
24 – 26	7
27 – 30	6
31 – 34	5
35 – 41	4
42 – 49	3
50 – 60	2
61 – 74	1
75 +	0

Age at first arrest

Ages at first arrest (8 to 85) are evaluated using the expression in the GRRS equation multiplied by 10, and rounded to the nearest whole number. 27 was subtracted from each risk contributions so the smallest contribution would be 0. For the 8 to 14 age group, the minimum risk contribution was chosen because so few individuals represent the youngest ages. Table 222 summarizes the results.

Table 222: Age at first arrest ranges and their contributions to the new raw risk score. Transformations are obtained by using the formula -0.613 $\times$ ln(age.first), multiplying the result by 10, selecting the median of the interval (or minimum for age at first arrest under age 14), and rounding to the nearest whole number. For ease of use, the risk contributions are shifted so that the smallest risk contribution is 0. This is accomplished by adding 27 to each possible risk contribution. 27 can be subtracted back when necessary for comparing the new raw scores to the original raw scores.

Age at first arrest	Risk contribution
8 – 14	11
15 – 17	10
18 – 20	9
21 – 24	8
25 – 28	7
29 – 33	6
34 – 39	5
40 – 46	4
47 – 54	3
55 – 64	2
65 – 75	1
76 +	0

Arrest rate

The transformation of the arrest rate calculation required the development of a matrix of outputs given relevant intervals covering 1) the number of years one has been involved in the criminal justice system (event period), and 2) a person's number of previous arrests. That matrix is show in Table 223. For all values within each combination of event period interval and number of previous arrests interval, the median value of the weighted arrest rates from the GRRS equation was obtained ($0.393 \times \ln(100 * \text{arrest.r})$). This was then multiplied by 10, rounded to the nearest whole number, and then 6 is subtracted from each risk contribution so that the smallest contribution is 0. Exceptions to this approach are:

- The arrest rate for people with no previous arrests was set at the median arrest rate for people in the current data who have at least 1 previous arrest, 0.55.
- The distribution of number of previous arrests greater than 18 is right skewed, so representative contributions were chosen from people with 19 to 25 previous arrests.
- The distribution of number of event periods greater than 26 years is right skewed, so representative contributions were chosen from people with 27 to 40 years in the criminal justice system.

Table 223: Risk contributions made by arrest rate. The risk contribution is the value of the cell in the row corresponding to the number of years in the criminal justice system (Age – Age at first arrest) and the column corresponding to the number of previous arrests.

		Number of previous arrests								
		0	1–2	3	4	5–6	7–8	9–12	13–18	19+
Years since first arrest	0	10	13	16	18	19	20	21	23	24
	1 – 2		10	13	14	15	16	18	19	21
	3 – 4		8	11	12	13	14	15	17	18
	5 – 6		6	9	10	11	13	14	16	17
	7 – 9		5	8	9	10	11	13	14	16
	10–12		4	7	8	9	10	12	13	14
	13–15		3	6	7	8	9	11	12	14
	16–20		2	5	6	7	8	10	11	13
	20–26		1	4	5	6	7	9	10	12
	27+		0	3	4	5	6	7	9	10

Note:

Years since first arrest =

Age at assessment (rounded to the age at nearest birthday) minus Age at first arrest.

The following three pages provide a graphical summary of the transformations of the Summative GRRS inputs.

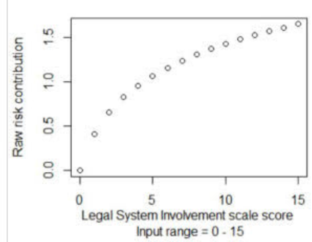
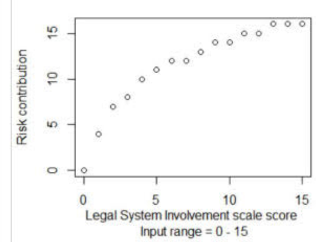
The COMPAS-R Summative General Recidivism Risk Scale (Summative GRRS) allows the determination of a person's recidivism risk level through simple addition, without involving complicated calculations. This tool was developed in response to users who wanted to understand and explain the way input changes affect a person's risk score. In a sense, the GRRS has always been summative because it has always been an additive model. However, calculating the values to add can seem daunting, and organizing all the possible results in a single scoring sheet would have been impossible because of the continuous nature of some of the variables.

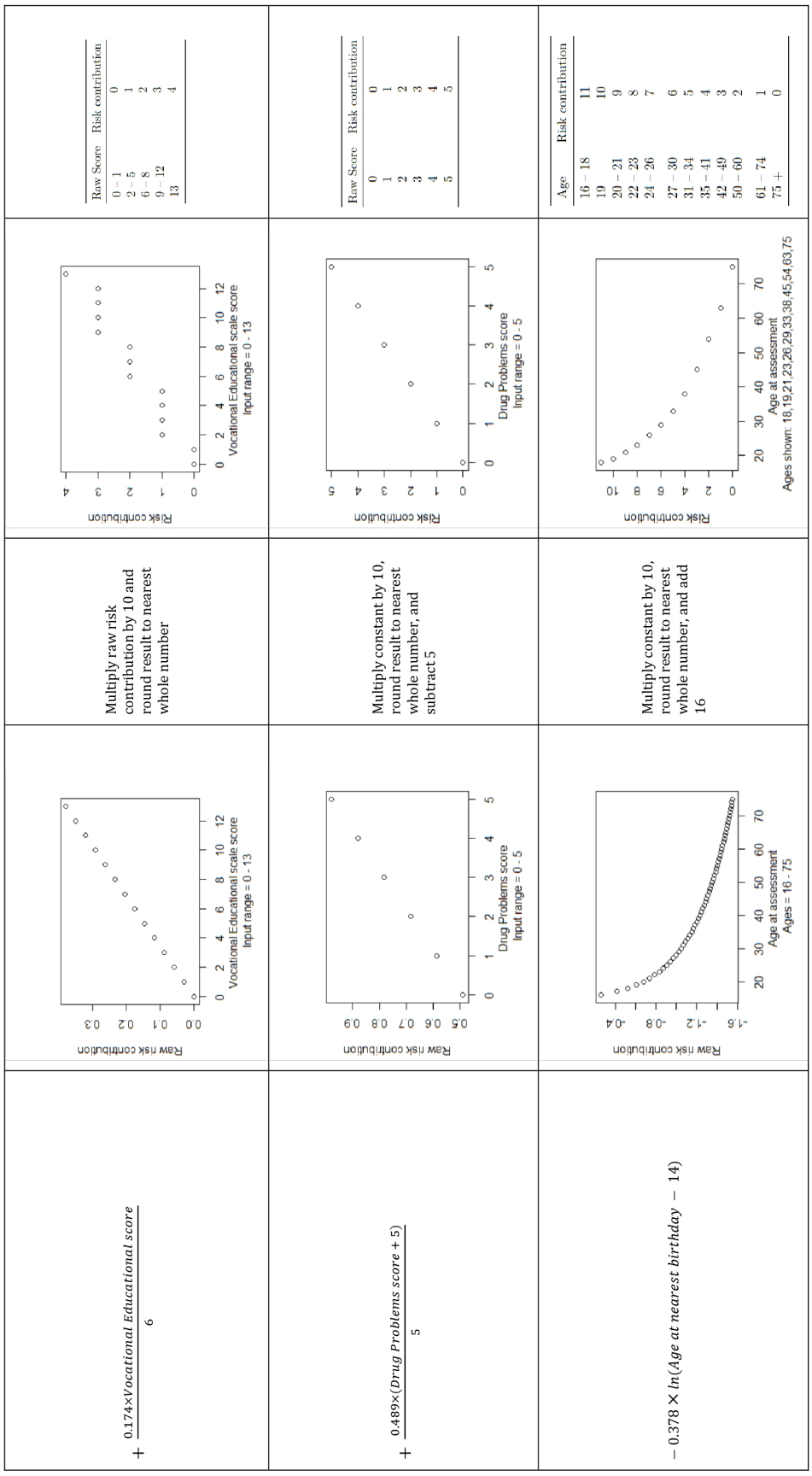
Logistic Regression Model
A new model was fit that borrowed from the COMPAS GRRS in structure. Six separate inputs are required in the model: (1) the Legal System Involvement scale score; (2) the Vocational Educational scale score; (3) the Drug Problems scale score; (4) age at first arrest; (5) age at assessment rounded to nearest birthday; (6) the number of previous arrests. The person's Raw Risk Score for the updated COMPAS-R Core General Recidivism Risk Scale is given by the following equation:

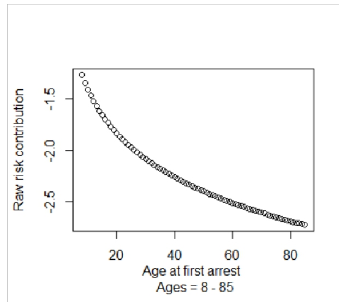
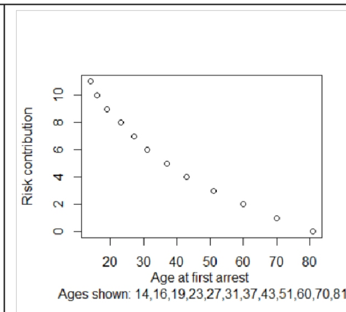
$$Raw\ Risk\ Score = -.683 + 0.593 \times \ln(Legal\ System\ Involvement\ score\ +\ 1) + \frac{0.174 \times Vocational\ Educational\ score}{6} + \frac{0.489 \times (Drug\ Problems\ score\ +\ 5)}{5}$$

$$-0.613 \times ln(Age\ at\ first\ arrest) - 0.378 \times ln(Age\ at\ nearest\ birthday\ -\ 14) + 0.393 \times \ln\left(\frac{100\ *\ (number\ of\ previous\ arrests + 1)}{Age\ at\ nearest\ birthday - Age\ at\ first\ arrest\ +\ 1}\right).$$

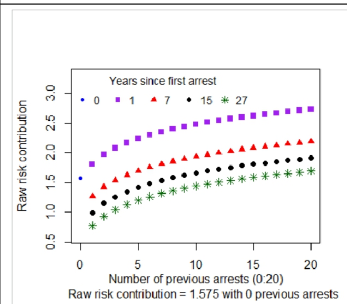
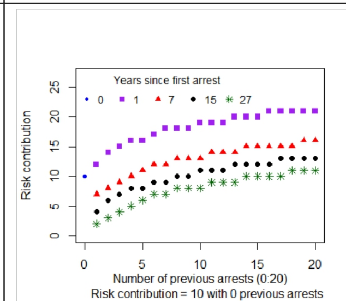
This document attempts to illustrate graphically the steps that were taken to transform the continuous inputs of the logistic model to discrete values that could be easily added together.

Term in Logistic Model: each term of the model is shown in this column.	**Raw Risk Contribution Plots:** the x-axis (horizontal axis) of each plot in this column shows the range of values the input can take. For example, the Legal System Involvement scale score can range from 0 to 15. The y-axis (vertical axis) shows the transformed values of the inputs expressed in that term of the model equation. The y-axis is labeled **Raw Risk Contribution** because that is the value that **would be added** to the Raw Risk Score from that term.	**Transformation to Whole numbers:** to transform each **Raw Risk Contribution** to a whole number that could be easily added, each Raw Risk Contribution was multiplied by 10, rounded to the nearest whole number, and the result adjusted by adding or subtracting a constant integer so that the smallest value would be 0. These values are called the **Risk Contributions**.	**Risk Contribution Plots:** As with the plots for the Raw Risk Contribution, the x-axis of each plot in this column shows the range of values that the input can take. The y-axis shows the Risk Contributions (whole number values) that result from the transformations described in the column to the left. The y-axis is labeled **Risk Contribution** because that is the value that **is added** to the **Total Risk Score** from that term.	**Table:** A more readable way to organize the values shown in the Risk Contribution Plots column.
−.683	Constant: no plot shown.	Multiply constant by 10, round result to nearest whole number, and add 7	No risk contribution.	Risk Contribution = 0
$+ 0.593 \times ln(Legal\ System\ Involvement\ score\ +\ 1)$		Multiply raw risk contribution by 10 and round result to nearest whole number		Raw Score — Risk contribution 0 — 0 1 — 4 2 — 7 3 — 8 4 — 10 5 — 11 6 – 7 — 12 8 — 13 9 – 10 — 14 11 – 12 — 15 13 – 15 — 16



<table>
<tr>
<td>$-0.613 \times ln(Age\ at\ first\ arrest)$</td>
<td> 
</td>
<td>Multiply constant by 10, round result to nearest whole number, and add 27.</td>
<td> 
</td>
<td>
<table>
<tr><th>Age at first arrest</th><th>Risk contribution</th></tr>
<tr><td>8 – 14</td><td>11</td></tr>
<tr><td>15 – 17</td><td>10</td></tr>
<tr><td>18 – 20</td><td>9</td></tr>
<tr><td>21 – 24</td><td>8</td></tr>
<tr><td>25 – 28</td><td>7</td></tr>
<tr><td>29 – 33</td><td>6</td></tr>
<tr><td>34 – 39</td><td>5</td></tr>
<tr><td>40 – 46</td><td>4</td></tr>
<tr><td>47 – 54</td><td>3</td></tr>
<tr><td>55 – 64</td><td>2</td></tr>
<tr><td>65 – 75</td><td>1</td></tr>
<tr><td>76 +</td><td>0</td></tr>
</table>
</td>
</tr>
<tr>
<td>$+0.393 \times \ln\left(\frac{100 * (number\ of\ previous\ arrests + 1)}{Age\ at\ nearest\ birthday - Age\ at\ first\ arrest + 1}\right)$

This term combines 2 inputs: the number of previous arrests and the number of years since first arrest (Age at nearest birthday – Age at first arrest). Here, we plot the risk contribution given four representatives “years since first arrest”: 1 year, 7 years, 15 years, and 27 years.

If a person had no previous arrests, their arrest rate was set at the median arrest rate for people who had at least 1 arrest, 0.55, yielding a raw risk contribution of 1.575: $0.393 \times \ln(100 \times 0.55) = 1.575$, and a transformed risk contribution of 10.</td>
<td> 
</td>
<td>Multiply constant by 10, round result to nearest whole number, and subtract 6.</td>
<td> 
</td>
<td>See Table 223 of the COMPAS-R Core Research Report or the Summative GRRS scoring table in the COMPAS-R Long Form to determine the Risk Contribution given years since first arrest and number of previous arrests.</td>
</tr>
<tr>
<td colspan="5">Notice that the sum of all of the adjustments is $7 - 5 + 16 + 27 - 6 = 39$. The overall effect of all transformations can be expressed as **10 × each input + 39**, with rounding error for each transformation that cannot be recovered when back-transforming the sum of the risk contributions. Additionally, to create a tool that would not be unwieldy to use, some input values in the Arrest Rate table that had similar risk impacts were grouped together (see Table 223 of the COMPAS-R Core Research Report or the Summative GRRS scoring table in the COMPAS-R Long Form).

Adding all Risk Contributions from each of the six terms yields a Total Risk Score. This can then be back-transformed by subtracting 39 and dividing the result by 10 to yield an *approximation* of the output by the logistic regression equation.</td>
</tr>
</table>

Mapping summative risk score back to new logistic risk score

To approximate the new logistic risk score reflected in the summative scale, subtract 32 from the total score, divide the result by 10 and subtract 0.683:

$$newGRRSraw \approx (GRRSsummative - 32)/10 - 0.683.$$

Table 224 summarizes the adjustments that need to be made to reverse the rounded scores. The overall adjustment can be made at once by subtracting 32 from the total score. Additionally, new scores were obtained by multiplying each term by 10: this is reversed in the formula by dividing by 10. Finally, the original equation has a constant that is subtracted from the sum of the original terms and we subtract that here as well.

Table 224: In the process of obtaining risk contributions from scales whose minimum risk contribution was not zero, we adjusted the contributions up or down by subtracting a constant from all the contributions of that scale. For example, the Drug Problems risk contributions went from 5 to 10. These were adjusted by subtracting 5 from each contribution to yield contributions of 0 to 5.

Input	Forward Adjustment	Back Adjustment
Legal System Involvement	no adjustment	no adjustment
Vocational/Educational	no adjustment	no adjustment
Drug Problems	–5	+5
Age at assessment	+16	–16
Age at first arrest	+27	–27
Arrest rate	–6	+6
Total score		–32

Table 225 gives an example of a person whose inputs were evaluated using both the new GRRS model (see Equation 1) and the COMPAS-R summative risk scale.

Table 225: The original GRRS takes the raw input values from the six sources shown in column 1 and uses Equation 1 to obtain the raw risk score; using the hypothetical values in column 2 (Raw value), this person obtains a GRRs raw score of –1.179. The COMPAS-R raw risk score is obtained by summing the risk contributions contained in this document; using the hypothetical values in column 2 (Raw value), this person obtains a COMPAS-R raw score of 33. To compare these scores, we subtract 32 from 33, divide the result by 10, and add in the constant, -0.683. The corresponding raw score is close, but because of rounding and using representative values from intervals, it is not expected that all scores will be this close.

Input	Raw value	New logistic GRRS contribution	COMPAS-R Risk contribution
Legal System Involvement	5	1.063	11
Vocational/Educational	13	0.377	4
Drug Problems	1	0.587	1
Age at assessment	43	-1.273	3
Age at first arrest	17	-1.737	10
Arrest rate	3 prev. arrests 43 – 17 = 26	1.059	4
Subtotal		= 0.076	= 33 – 32 = 1 /10 = 0.1
Constant		– 0.683	– 0.683
Raw score for the original GRRS		= – 0.607	= – 0.583

11 Bibliography

American Psychiatric Association, and others. 2013. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. American Psychiatric Pub.

Andrews, D. A., and James Bonta. 2016. *The Psychology of Criminal Conduct, 6th Edition*. Routledge.

Box, George E. P. 1979. "Some Problems of Statistics and Everyday Life." *Journal of the American Statistical Association* 74 (365): 1–4.

Brennan, Tim, William Dieterich, and William Oliver. 2007. *COMPAS Technical Manual and Psychometrics Report Version 5.01*. Traverse City, MI: Northpointe Institute for Public Management.

Brewin, Chris R., Suzanna Rose, Bernice Andrews, John Green, Philip Tata, Chris McEvedy, Stuart Turner, and Edna B. Foa. 2002. "Brief Screening Instrument for Post-Traumatic Stress Disorder." *The British Journal of Psychiatry* 181 (2): 158–62.

Cohen, Lawrence E., and Marcus Felson. 1979. "Social Change and Crime Rate Trends: A Routine Activity Approach." *American Sociological Review* 44: 588–608.

Coleman, J. S. 1990. *Foundations of Social Theory*. Cambridge,Mass: Harvard University Press.

Compton, Elizabeth, Michael Bentley, Sharon Ennis, and Sonya Rastogi. 2010. "2010 Census Race and Hispanic Origin Alternative Questionnaire Experiment." *2010 Census Program for Evaluations and Experiments*.

Connor, Kathryn M, and JRT Davidson. 2001. "SPRINT: A Brief Global Assessment of Post-Traumatic Stress Disorder." *International Clinical Psychopharmacology* 16 (5): 279–84.

Cooke, D. J., Adelle E. Forth, and R. D. Hare, eds. 1998. *Psychopathy: Theory, Research and Implications for Society*. Dordecht, Netherlands: Kluwer Academic Publishers.

Cortina, J. M. 1993. "What Is Coefficient Alpha? An Examination of Theory and Applications." *Journal of Applied Psychology* 78: 98–104.

Davidson, Jonathan. 2002. "SPAN Addendum to DTS Manual." *New York: Multi-Health System Inc*.

DeVellis, Robert F. 2016. *Scale Development: Theory and Applications*. Vol. 26. Sage Publications.

Fabrigar, L. R., D. T. Wegener, R. C. MacCallum, and E. J. Strahan. 1999. "Evaluating the Use of Factor Analysis in Psychological Research." *Psychological Methods* 4: 272–99.

Farrington, D. P. 1991. "Childhood Aggression and Adult Violence: Early Precursors and Later Life Outcomes." In *The Development and Treatment of Childhood Aggression*, edited by D. J. Pepler and K. H. Rubin, 5–29. Hillsdale, NJ: Lawrence Erlbaum.

Gendreau, P., Goggin C., and T. Little. 1996. "A Meta-Analysis of the Predictors of Adult Offender Recidivism: What Works!" *Criminology* 34: 575–607.

Gottfredson, M. R., and T. Hirschi. 1990. *A General Theory of Crime*. Stanford: CA: Stanford University Press.

Graham, J. R. 2000. *MMPI-2: Assessing Personality and Psychopathology*. 3rd ed. New York: Oxford University Press.

Grayson, DA. 1988. "Two-Group Classification in Latent Trait Theory: Scores with Monotone Likelihood Ratio." *Psychometrika* 53 (3): 383–92.

Grove, W. M., D. H. Zald, B. S. Lebow, B. E. Snitz, and C. Nelson. 2000. "Clinical Versus Mechanical Prediction: A Meta-Analysis." *Psychological Assessment* 12 (1): 19–30.

Hagan, J. 1998. "Life Course Capitalization and Adolescent Development." In *New Perspectives on Adolescent Risk Behavior*, edited by Richard Jessor, 499–517. New York: Cambridge University Press.

Hanson, Robert Karl, Guy Bourgon, Robert J. McGrath, Daryl Kroner, David A. D'Amora, Shenique S. Thomas, and Lahiz P. Tavarez. 2017. *A Five-Level Risk and Needs System: Maximizing Assessment Results in Corrections Through the Development of a Common Language*. WS Hein. https://csgjusticecenter.org/wp-content/uploads/2020/01/A-Five-Level-Risk-and-Needs-System_Report.pdf.

Hare, R. D. 1991. *Manual for the Hare Psychopathy Checklist-Revised*. Toronto, Canada: Multi-Health Systems.

Hemker, Bas T, Klaas Sijtsma, Ivo W Molenaar, and Brian W Junker. 1996. "Polytomous IRT Models and Monotone Likelihood Ratio of the Total Score." *Psychometrika* 61 (4): 679–93.

Hirschi, Travis. 1969. *Causes of Delinquency*. Berkeley, CA: University of California Press.

Hoffmann, Norman G, Dana E Hunt, William M Rhodes, and K Jack Riley. 2003. "UNCOPE: A Brief Substance Dependence Screen for Use with Arrestees." *Journal of Drug Issues* 33 (1): 29–44.

Huynh, Huynh. 1994. "A New Proof for Monotone Likelihood Ratio for the Sum of Independent Bernoulli Random Variables." *Psychometrika* 59 (1): 77–79.

Kaeble, Danielle, and Mary Cowhig. 2018. "Correctional Populations in the United States, 2016 (NCJ 251211)." *Washington, DC: US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics*.

Lykken, D. 1995. *The Antisocial Personalities*. Hillsdale, N.J.: Lawrence Erlbaum.

Marlowe, Douglas B. 2018. "The Most Carefully Studied, yet Least Understood, Terms in the Criminal Justice Lexicon: Risk, Need, and Responsivity." https://www.prainc.com/risk-need-responsitivity/.

Moffitt, Terrie E. 1993. "Adolescence-Limited and Life-Course Persistent Antisocial Behavior: A Developmental Taxonomy." *Psychological Review* 100: 674–701.

Northpointe Inc. 2019. "Practitioner's Guide to COMPAS Core." Traverse City, MI. 2019. http://www.equivant.com/wp-content/uploads/Practitioners-Guide-to-COMPAS-Core-040419.pdf.

———. 2021. "Practitioner's Guide to COMPAS-R Core." Traverse City, MI. 2021.

Osgood, D. W., J. K. Wilson, P. M. O'Malley, J. G. Bachman, and L. D. Johnston. 1996. "Routine Activities and Individual Deviant Behavior." *American Sociological Review* 61 (4): 635–55.

Parker, J. G., and S. R. Asher. 1987. "Peer Relations and Later Personal Adjustment: Are Low-Accepted Children at Risk." *Psychological Bulletin* 102 (3): 357–89.

Prins, Annabel, Michelle J. Bovin, Derek J. Smolenski, Brian P. Marx, Rachel Kimerling, Michael A. Jenkins-Guarnieri, Danny G. Kaloupek, et al. 2016. "The Primary Care PTSD Screen for DSM-5 (PC-PTSD-5): Development and Evaluation Within a Veteran Primary Care Sample." *Journal of General Internal Medicine* 31 (10): 1206–11.

Quinsey, V. L., G. T. Harris, M. E. Rice, and C. A. Cormier. 1998. *Violent Offenders: Appraising and Managing Risk*. Washington, DC: American Psychological Association.

Robin, Xavier, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez, and Markus Müller. 2011. "pROC: An Open-Source Package for r and s+ to Analyze and Compare ROC Curves." *BMC Bioinformatics* 12: 77.

Sullivan, Lisa M., Joseph M. Massaro, and Ralph B. D'Agostino Sr. 2004. "Presentation of Multivariate Data for Clinical Use: The Framingham Study Risk Score Functions." *Statistics in Medicine* 23 (10): 1631–60.

Swets, J. A., R. M. Dawes, and J. Monahan. 2000. "Psychological Science Can Improve Diagnostic Decisions." *Psychological Science in the Public Interest* 1: 1–26.

Van der Ark, L Andries, and others. 2007. "Mokken Scale Analysis in r." *Journal of Statistical Software* 20 (11): 1–19.

———. 2012. "New Developments in Mokken Scale Analysis in r." *Journal of Statistical Software* 48 (5): 1–27.

Wieland, Andreas, Christian F. Durach, Joakim Kembro, and Horst Treiblmaier. 2017. "Statistical and Judgmental Criteria for Scale Purification." *Supply Chain Management: An International Journal*.