The Concurrent and Predictive Validity of a Needs and Responsivity Assessment System

Authors
Grant Duwe, Ph.D.
Research Director
Email: grant.duwe@state.mn.us

Valerie Clark, Ph.D.
Research Supervisor
Email: valerie.clark@state.mn.us

1450 Energy Park Drive, Suite 200
St. Paul, Minnesota 55108-5219
651/361-7200
TTY 800/627-3529
www.doc.state.mn.us
December 2023

Research Summary

Using a sample of nearly 2,100 people incarcerated in Minnesota's prison system, this study examined the concurrent and predictive validity of a needs and responsivity assessment system. For concurrent validity, we evaluated the relationship between the 13 needs and responsivity domains with assessed recidivism risk levels. For predictive validity, we analyzed the association between the domains and recidivism for a sub-sample that had been released from prison prior to 2023. The hypothesized needs domains—anti-social thinking, anti-social peers, education, employment, substance use disorder, housing/homelessness, and family/domestic—were significantly associated with assessed and observed recidivism, while most of the hypothesized responsivity domains—mental health, religiosity, motivation and learning style—were not. The results suggest self-identity is a distinct criminogenic need. Gender and racial/ethnic differences for concurrent and predictive validity were relatively minimal across the 13 domains.

The risk-needs-responsivity (RNR) model has become the prevailing paradigm used to guide the delivery of programming to correctional populations. According to the RNR model, programming should be calibrated to an individual's risk of reoffending, criminogenic needs, and responsivity issues (Gendreau, French, & Gionet, 2004). Whereas the risk principle holds that interventions should be targeted toward higher-risk individuals, the needs principle suggests that programs must address individual characteristics that are related to criminal behavior; i.e., criminogenic needs (Sperber, Latessa, & Makarios, 2013). And the responsivity principle dictates that programming should be tailored to a person's strengths, learning style, and abilities.

Assessments are central to the RNR model, for they help determine a person's risk for recidivism (or other harmful outcomes such as prison misconduct), their criminogenic needs areas that should be addressed through programming, and specific responsivity factors that may influence whether they successfully complete an intervention (Bonta & Andrews, 2017). Currently, a wide variety of assessment instruments are available for correctional populations that were designed to assess their risk for recidivism and prison misconduct, their criminogenic needs, and their responsivity issues. Some instruments have been created to simultaneously assess for risk, needs and responsivity, whereas others focus only on assessing recidivism risk or specific domains for criminogenic needs (e.g., criminal thinking) and responsivity (e.g., childhood trauma).

Regardless of whether an instrument was designed to assess a specific domain or all three areas of the RNR model, it is critical to evaluate its validity. Concurrent validity and predictive validity are two components of criterion validity, which examines how an assessment effectively estimates an individual's performance on an outcome measure (i.e.,

criterion). If an assessment can accurately predict the outcome measure(s), then the instrument has criterion validity. Both concurrent and predictive validity examine the relationship between the assessment and the outcome measure, but there is a difference in the time the criteria (i.e., outcomes) are measured (McIntire & Miller, 2005). In concurrent validity, the assessment result and the criterion variable are measured simultaneously. With predictive validity, however, the criterion variable is measured at some point after the assessment was administered.

Present Study

In this study, we evaluate the concurrent and predictive validity of a needs and responsivity assessment system that was piloted within the Minnesota Department of Corrections (MnDOC) in 2021. The risk component of this system is the Minnesota Screening Tool Assessing Recidivism Risk (MnSTARR) 2.0, an instrument that has been validated on Minnesota's prison population (Duwe, 2021). The needs and responsivity components consist of assessments for 13 domains: 1) anti-social thinking, 2) anti-social peers, 3) substance use disorder (SUD), 4) education, 5) employment, 6) housing stability, 7) family/domestic relationships, 8) religious faith and spirituality, 9) mental health, 10) childhood trauma, 11) self-identity, 12) motivation and 13) learning style. Of the 13 needs and responsivity domains, the MnDOC has existing assessment processes for SUD, education, and mental health. For the remaining 10 domains, we developed a self-reported assessment that was administered in March 2021 to individuals confined in Minnesota prisons.

Our sample for this study consists of nearly 2,100 people who completed the self-reported assessment. To examine concurrent validity, we analyze the degree to which these

13 domains are associated with MnSTARR 2.0 risk levels. For predictive validity, we evaluate the relationship of the 13 domains with recidivism for a sub-sample of individuals who had been released from prison prior to January 2023. For both the concurrent and predictive validity analyses, we also examine whether the results vary by gender and race/ethnicity.

RNR Assessments for Correctional Populations

Prior research suggests that predictors of criminal behavior (i.e., criminogenic needs) can be grouped within eight main domains. Within what has come to be known as the "central eight", recidivism risk factors have been characterized as major and moderate. The four major risk factors, which have been referred to as the "Big Four", include history of antisocial behavior (i.e., criminal history), antisocial personality pattern, antisocial cognition, and antisocial associates. The four moderate risk factors, on the other hand, include family/marital relationships, education/employment, leisure/recreation, and substance use (Andrews, Bonta, and Wormith, 2006). Although criminal history is typically the strongest predictor of future criminal behavior (Caudy, Durso, & Taxman, 2013), it is a static factor that cannot be changed through interventions. The seven remaining criminogenic needs are dynamic risk factors that can be targeted through interventions because changes can be made in these factors.

The responsivity principle consists of two parts—general and specific. Whereas general responsivity refers to types of programming that are most effective in reducing recidivism, such as cognitive-behavioral interventions, specific responsivity includes individual factors that may influence successful involvement in programming (Bonta and Andrews, 2017). Examples of specific responsivity factors include motivation, anxiety,

learning style, mental health, religiosity, language, gender, and culture (Cullen, 2002; McCormick et al., 2017; Mowen et al., 2018; Pinals et al., 2021; Sachs & Miller, 2018).

The Development of the MnDOC RNR Assessment System

The impetus for the MnDOC developing a comprehensive RNR assessment system arose from perceived limitations with the agency's existing assessment process. In the early 2000s, the MnDOC began using the Level of Service Inventory-Revised (LSI-R). Beginning in 2013, the agency transitioned to the Level of Service/Case Management Inventory (LS/CMI), but the agency used it only as a needs assessment instrument due to the debut of the original MnSTARR that same year (Duwe, 2014). That is, the MnDOC used the MnSTARR to assess recidivism risk and the LS/CMI to assess the needs for people identified as higher risk on the MnSTARR. In addition to these instruments, the agency administered assessments for substance use disorders, education, and physical and mental health.

After using the MnSTARR, which had been manually scored by prison caseworkers, for three years, the MnDOC shifted to the MnSTARR 2.0 in November 2016 (Duwe & Rocque, 2017). The MnSTARR 2.0 is a gender-specific, fully-automated recidivism risk assessment instrument that is run on everyone who enters Minnesota's prison system, regardless of their length of stay. As shown in a recent revalidation study, the MnSTARR 2.0 accurately predicts recidivism for Minnesota's prison population, achieving an overall area under the curve (AUC) of 0.74 for females and 0.73 for males (Duwe, 2021). The MnSTARR 2.0 contains nearly 50 items to predict multiple types of recidivism[1], which are then used to assign individuals to one of four risk levels: 1) Low, 2) Medium, 3) High, and 4)

[1] The MnSTARR 2.0 includes items pertaining to criminal history (severity and type of offense), demographic characteristics (age at release and marital status), index offense type, prison admission type, prison misconduct, gang affiliation, and successful participation in effective MnDOC interventions such as substance use disorder treatment, earning education degrees and prison visitation.

Very High. The Very High category contains the top 20 percent of the highest risk individuals, the High category the next 20 percent, the Medium category the middle 20 percent, and the Low category the bottom 40 percent.

Similar to the MnSTARR 2.0, which has yielded recidivism risk assessments for everyone in Minnesota's prison system, the vast majority of individuals entering prison receive physical and mental health screenings. The other assessments, however, have been administered more selectively. For example, according to MnDOC policy, SUD and education assessments are administered only to people who have 150 days or more to serve in prison. Moreover, due to the length of time (approximately 60-90 minutes) it takes caseworkers to complete a LS/CMI, this assessment was administered only on high-risk individuals (i.e., top 40 percent per the MnSTARR 2.0) who have six months or more to serve in prison. Nevertheless, nearly half of the individuals released from Minnesota prisons from 2016-2017 who should have been scored on the LS/CMI did not receive this assessment (Duwe, 2021). Even among those who were scored on the LS/CMI, MnDOC staff were not consistently using the assessments to develop case plans and identify appropriate programming.

Amid the observed deficiencies with its overall assessment process, the MnDOC decided to develop and pilot an assessment system that delivers a more comprehensive measurement of needs and responsivity for a broader swath of the MnDOC population. To this end, the MnDOC retained the MnSTARR 2.0 and its SUD, education, and physical and mental health assessment processes. Based on the SUD assessment, individuals receive a rating ("High", "Medium", or "Low") that signifies the need to address that area. For education, individuals were given a need rating based on their level of achievement, with

"High" assigned to those with less than a secondary degree, "Medium" given to those with a secondary degree, and "Low" assigned to those with a post-secondary degree or certificate. Based on a mental health screening administered at intake, individuals receive a mental health score, ranging from 0 to 7, in which points are assigned for the presence of mental health disorders (e.g., mood disorder) and psychological concerns such as suicidal tendencies and a history of self-injury.

The self-reported assessment encompasses the following 10 domains: 1) Anti-Social Thinking, 2) Anti-Social Peers, 3) Self-Identity, 4) Childhood Trauma, 5) Employment, 6) Housing and Homelessness, 7) Family/Domestic, 8) Religiosity, 9) Motivation and 10) Learning Style. To the extent possible, the self-reported assessment relied on scales that have been validated or used in prior research. Accordingly, the Texas Christian University-Criminal Thinking Scales (TCU-CTS version 2) were used to measure Anti-Social Thinking (Knight et al., 2006), the Attitudes towards Associates scale within the Measures of Criminal Attitudes and Associates (MCAA) was used to measure Anti-Social Peers (Mills, Kroner, & Forth, 2002), the Multidimensional Scale of Perceived Social Support (MSPSS) was used to measure relationships with family, significant others, and peers (Zimet et al., 1988), the adverse childhood experiences (ACEs) scale was used to measure Childhood Trauma (Felitti et al., 1998), and the Visual, Auditory, Reading, Kinesthetic (VARK) instrument was used to assess Learning Style (Fleming & Mills, 1992).

To measure Self-Identity, the self-reported assessment included two measures derived from the literature. The first, adapted from Paternoster and Bushway's (2009) "feared self" concept, included six items such as "All in all I feel like I am a failure" and "I have hit rock bottom in my life." The second identity measure was based on Giordano and colleagues'

(2002) concept of cognitive transformation and included six items related to how open individuals are to change and whether they want to choose a different path. The cognitive change measure included items such as “I am open to a new way of life” and “I want to avoid criminal behaviors.”

The Religiosity scale drew from three items that have been used in prior research with correctional populations (Jang et al., 2017). Questions for this scale included: 1) In general, how important is religion to you?, 2) How often do you prefer to attend religious services?, and 3) “About how often do you spend time alone praying or reading the Bible, Koran, Torah, or other sacred book?”.

Due to the lack of existing validated assessments for Employment, Housing/ Homelessness and Motivation, MnDOC staff created measures for each domain. The Employment scale asked individuals about work history, career planning, and job search skills. A higher score for this domain suggests it is a high need due to limited employment experience and job skills. The Housing/Homelessness scale asked respondents about prior experiences with homelessness, the frequency and duration of prior homeless episodes in the past, and the anticipated living situation upon release. Again, a higher score for this measure reflects a history of homelessness and anticipated housing instability following release from prison. Finally, the Motivation scale contains items that asked individuals to identify the top two areas in which they need to make a change to avoid future involvement in the criminal justice system and their level of confidence (values ranged from 0 to 10) in being able to make those changes. For example, if a person identified sobriety and peer influences as their top two areas and gave confidence values of 9 for each one, the average Motivation score (9 out of 10) reflects a relatively high degree of confidence in making a change.

Assessing Construct Validity

Because needs and responsivity domains can be theoretical (or latent) in nature, testing for construct validity is important to determine whether domains are adequately scaled and working collectively to measure constructs such as anger/hostility, antisocial attitudes/behaviors, and cognitions/skills (Mei et al., 2021). Construct validity consists of five key components: 1) content, 2) convergent/divergent, 3) internal (latent structure), 4) concurrent and 5) predictive validity.

Content validity is the degree to which the items are relevant to, and representative of, the defined construct. Moreover, it is typically a measure of agreement between raters, who are considered content experts. The MnDOC's needs and responsivity assessment system was developed on the basis of direct input from staff whom the agency considers to be subject matter experts. Accordingly, the assessment system has content validity.

Convergent and divergent validity assess the extent to which needs items are converging or diverging within and across domains. More specifically, whereas convergent validity tests that constructs that are expected to be related are, in fact, related, divergent validity tests that constructs that should have no relationship do, in fact, not have any relationship. The present study assesses concurrent and predictive validity, while a separate study used the same data to conduct exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) to test convergent/divergent validity (Clark and Duwe, 2023. The results of this study are described in the next section.

Revisions to the Needs and Responsivity Assessment System

The results from the EFA and CFA, which assess the strength of items' loadings and cross-loadings across needs and responsivity domains, showed that most of the domains had

a high level of consistency (Clark and Duwe, 2023). This finding may reflect the fact that the needs and responsivity assessment system relies, to a large extent, on validated assessment scales such as the TCU-CTS, the MCAA, ACEs, and the MSPSS. Nevertheless, the EFA and CFA results identified several adjustments that needed to be made to the needs and responsivity assessments.

While no adjustments were made to the TCU-CTS for the present study, the EFA/CFA results reported by Clark and Duwe (2023) were consistent with the analyses conducted by Taxman et al. (2011). That is, the TCU-CTS demonstrated acceptable psychometrics with a six-factor solution, with factor loadings ranging from 0.483 to 0.827. However, three of the six subscales had weak discriminate validity. The Entitlement, Justification, and Personal Irresponsibility scales were strongly correlated—though not as strongly correlated as they were in Taxman et al.'s (2011) analysis—and had significant cross-loadings. The TCU-CTS was recently revised to remove stigmatizing language, items with poor face validity, and one sub-index that trivialized racially-based cynicism towards the criminal justice system (Sease and Knight, 2022). Initial testing indicates that this new version of the TCU-CTS has improved validity and will replace the previous version in subsequent iterations of the self-reported needs and responsivity assessment. As shown later in Table 1, the overall TCU-CTS scores ranged from 6 to 43 within our sample.

The Self-Identity measures had acceptable fit statistics after some adjustments were made. The EFA found that the feared self measure (Paternoster and Bushway, 2009) had optimal psychometric properties with a two-factor solution. The following three items that measure the respondent's realization of the consequences of deviant behavior loaded on to one factor: (1) "I worry that I will eventually burn out if I stay on the same path.", (2) "I have

hit rock bottom in my life.", and (3) "I need to make a change for the better to become more satisfied in my life." The second factor tapped into whether the respondent was happy with the direction of their life using the following three items: (1) "I am headed in the right direction in life.", (2) "I like the person I have become.", and (3) "I am becoming a better person." The EFA found that only three of the original six items included in the cognitive transformation scale (Giordano et al., 2002) loaded on to one factor. These items included: (1) "It is important to me that family and friends think of me as a good person.", (2) "I am open to a new way of life.", and (3) "I can envision a new life for myself." The CFA produced a solution that left the first factor of the feared self scale as a stand-alone construct (factor loadings ranging from 0.597 to 0.685), and the second factor of the feared self scale paired with the three-item cognitive transformation scale as a two-factor construct (factor loadings ranging from 0.538 to 0.969). The overall Self-Identity score ranged from 4 to 18 for our sample, with higher scores reflecting a more anti-social identity.

The self-reported needs and responsivity assessment originally included 16 items to measure the Employment domain. However, the EFA and CFA found that only half of these items loaded on to two factors. The first factor included items that relate to the respondent's work history. These items included whether the respondent: (1) was employed in the year prior to incarceration, (2) has a lot of work experience, (3) has prior employers that would provide a favorable job reference, and (4) liked the jobs they have had in the past. The second factor assessed the respondent's career future, including whether the respond knows (1) about jobs in his/her home community, (2) what kind of job he or she wants when released from prison, (3) what his or her career future looks like, and (4) how to explain gaps in his or her work history. The CFA found good fit statistics for the Employment domain,

although there were some significant cross-loadings between the two factors and the factors were strongly correlated. The factor loadings ranged from 0.433 to 0.939. The Employment score ranged from 15 to 67 for our sample, with higher scores signifying a lack of prior work experience and a limited capacity for finding a job in the future.

The Family/Domestic domain was originally measured using seven items, including relationship status, support from family, and relationships with children. However, the EFA revealed that only the four items relating to children loaded onto a single factor (quality of relationship with children, whether children lived with them prior to incarceration, having active child support orders, and whether parental rights have been terminated). Based on the CFA, factor loadings for this construct ranged from 0.483 to 0.826. To round out the Family/Domestic domain with measures of support from family and significant others, two of the three sub-indices included in the MSPSS were used. The CFA found that the family and significant other sub-indices of the MSPSS had factor loadings ranging from 0.917 to 0.956. Overall, the Anti-Social Peers score ranged from 2 to 50 for our sample, with higher scores reflecting stronger associations with friends and acquaintance involved in crime.

The third sub-scale of the MSPSS (support from friends) was used to supplement the Anti-Social Peers domain. The four-item friends sub-scale from the MSPSS had factor loadings ranging from 0.907 to 0.949. The ten-item MCAA (Attitudes Towards Associates) scale loaded strongly on to one factor with good fit statistics (factor loadings ranged from 0.547 to 0.903).

To assess the Mental Health domain, the self-reported needs and responsivity assessment originally included eight items derived from the mental health screening that occurs at intake. The EFA and CFA found that only four of the original eight items loaded on

to one factor with acceptable fit statistics. These items included whether the respondent had a mental illness or mood disorder, and whether the respondent was at risk of suicide or self-injury. Factor loadings for this domain ranged from 0.490 to 0.972. Scores for the Mental Health domain ranged from 0 to 7, with higher scores signifying multiple mental health indicators.

The remaining two domains that were analyzed by Clark and Duwe (2023)—Religiosity and Housing/Homelessness—all demonstrated optimal psychometric properties and required no adjustments. The scores for Religiosity ranged from 0 to 18, with higher scores measuring greater religious faith and involvement. Housing/Homelessness scores ranged from 0 to 10, with higher scores denoting more prior experiences with homelessness and greater anticipated housing instability following release from prison. Our evaluation of concurrent and predictive validity is based on the revised version of the needs and responsivity assessments. While some domains may contain multiple factors, only the overall scores are used in the present study's analyses.

Four domains—SUD, Education, Motivation, and Learning Style—were not included in Clark and Duwe (2023). The SUD domain was not included because the researchers did not have access to the assessment data used to make the determination of high, medium, or low need for treatment. Given that three indicators are required to measure a latent construct, the Education and Motivation domains were not included because they were based on one or two indicators. The Learning Style domain (the VARK) was not included because it was primarily intended to inform programming staff on educational preferences, and it was validated in a previous study (Leite et al., 2010).

The Empirical Distinction between Needs and Responsivity

A key distinction between criminogenic needs and responsivity factors lies in the impact each one has on recidivism. A criminogenic need will not only have a significant, direct impact on criminal behavior, which includes outcome measures like recidivism, but interventions that successfully target this need will reduce reoffending. For example, substance use is a criminogenic need with a significant, direct impact on recidivism (Gendreau et al., 1996), and substance use disorder treatment has been shown to reduce reoffending (Mitchell et al., 2007).

Specific responsivity factors, on the other hand, are not expected to have a significant, direct impact on criminal behavior. Instead, specific responsivity factors will have a more modest, indirect influence on reoffending that is moderated by other factors. Moreover, as McCormick et al. (2017) indicate, responsivity factors may affect whether individuals are able to successfully complete programming that targets criminogenic needs.

At least seven of the 13 domains included within the MnDOC's needs and responsivity assessment are widely considered to be criminogenic needs. For example, the existing literature has identified anti-social thinking, anti-social peers, substance use disorders, education, employment, and family/domestic relationships as criminogenic needs (Andrews et al., 2006). Moreover, existing research has generally found that housing instability is associated with recidivism (Clark, 2016; Kirk et al., 2018; Steiner et al., 2015). As such, we hypothesize these seven domains will have statistically significant results in the concurrent and predictive validity analyses.

Among the remaining six domains, the literature has consistently identified learning style and motivation as specific responsivity factors (Sachs & Miller, 2018). But for some of the other domains, the extant literature has not always provided clear-cut guidance as to

which category—needs or responsivity—each one belongs. Consider, for example, what the literature has indicated about mental health. Andrews and colleagues (2006) acknowledged that while mental illness is a minor risk factor for recidivism, they emphasized it has only a modest, indirect impact on reoffending. Whatever effect mental illness has on recidivism, Andrews et al. (2006) argued, likely reflects the impact of substance use (one of the "central eight" risk factors) along with criminal thinking and antisocial personality pattern (two of the "big four"). Other scholars, however, have identified mental health as a specific responsivity factor that may moderate the success of interventions targeted to criminogenic needs (McCormick et al., 2017; Pinals et al., 2021).

Similar points have been made about religiosity, with some research indicating it has a significant, albeit modest, effect on crime and delinquency (Johnson, De Li, Larson, and McCullough, 2000). Other research, however, posits that religious faith and spirituality should be considered as a type of specific responsivity, given that individuals who identify in religious or spiritual ways have been found to experience the positive effects of prosocial support (Mowen, Stansfied, and Boman, 2018). We hypothesize that both mental health and religiosity will be specific responsivity factors insofar as they will *not* have a significant, direct impact on assessed and observed recidivism. Instead, consistent with how specific responsivity factors are conceptualized, we anticipate that each one would significantly influence successful involvement in programming. However, validating the specific responsivity factors with program participation and outcome data is beyond the scope of this study.

While research suggests that self-identity and, more specifically, identity transformation is crucial to the desistance process (Giordano et al., 2002; Paternoster &

Bushway, 2009; Rocque et al., 2016), the literature has had relatively little to say as to whether it should be considered a criminogenic need or a specific responsivity factor. The same could be said about childhood trauma, which is a static factor for an adult correctional population. Without strong evidence indicating that either one has a significant, direct impact on recidivism, we hypothesize that self-identity and childhood trauma will be specific responsivity factors.

Data and Method

The needs and responsivity assessment was self-administered on desktop computers using Snap computer-assisted survey software. Incarcerated individuals selected to participate in the pilot were notified in writing by their case manager about one week prior to taking the assessment. Individuals were advised that their participation in the pilot was completely voluntary and they could refuse to participate or skip any questions that they did not want to answer. Incarcerated individuals signed a consent form prior to beginning the pilot, and respondents were offered a small incentive in exchange for their participation.

The self-reported assessment was administered to incarcerated individuals at all 11 adult prisons in Minnesota in the spring of 2021. These facilities include a range of custody levels from minimum to maximum throughout the state. All but one of the facilities house men, while the remaining facility houses women. In an effort to achieve the largest sample possible without unduly burdening staff at the men's facilities, half of the approximately 6,700 men who were incarcerated at the time of the survey were randomly selected. Given the relatively small number of incarcerated women (400), all individuals housed in the lone women's facility were invited to participate. Of the 3,335 men and 400 women who were invited to participate, a total of 2,117 agreed to take the assessment. After removing 22

individuals who did not provide responses to more than 5% of the items, our final sample consisted of 1,758 men and 337 women, resulting in a total participation rate of 56 percent (53 percent for men, and 84 percent for women).

Measures and Analytic Strategy

While predictive validity provides a more direct and, thus, more meaningful evaluation of the 13 needs and responsivity domains, it also important to consider concurrent validity. Of the 2.095 individuals who completed the self-reported assessment, a little more than half had been released by the end of 2022. Therefore, data on recidivism, which we use to assess predictive validity, were unavailable for almost half of our sample. On the other hand, data on MnSTARR 2.0 assessments, which we use to measure concurrent validity, were available for all 2,095 respondents. Because the MnSTARR 2.0 is a fully-automated instrument, every individual admitted to prison is assessed at least once and, in most instances, receives three assessments prior to release.[2]

For concurrent validity, we evaluated the relationship between the 13 domains and the MnSTARR 2.0 risk level from the most recent assessment prior to the completion of the self-reported assessment. More specifically, we computed correlation coefficients and the area under the curve (AUC) statistics. When individuals are assessed on the MnSTARR 2.0, they are assigned to one of four possible risk levels. To generate AUC values, we collapsed the "high" and "very high" levels into one category (value = 1) and the "low" and "medium" levels into a second category (value = 0). For these analyses, the AUC statistic is interpreted as the probability that a randomly selected individual with a rating of "high" or "very high"

[2] MnSTARR 2.0 assessments are automatically generated on 1) the day a person enters prison, 2) every 365 days a person is in prison, 3) 130 days prior to release for release planning purposes and 4) the day of release. MnDOC staff can also run assessments on their own at any time for individuals who are incarcerated.

on the MnSTARR 2.0 has a higher domain score than a randomly selected individual with a rating of "medium" or "low". AUC values at either end of the spectrum (0 or 1) reflect perfect prediction, whereas a value of 0.50 indicates an assessment does no better than chance.

For predictive validity, we also calculated correlation coefficients and AUC values between the 13 domains and recidivism for a sub-sample of respondents who were released from prison by the end of 2022. For these analyses, the AUC statistic is interpreted as the probability that a randomly selected recidivist has a higher domain score than a randomly selected non-recidivist. We measured recidivism as a reconviction for a new offense that occurred between the time of release from prison and June 30, 2023. Data on reconvictions, which we used to create a binary measure (1 = reconviction, 0 = no reconviction), were obtained from the Minnesota Bureau of Criminal Apprehension.

We further evaluated concurrent and predictive validity by comparing the performance of the 13 domains across gender and race/ethnicity. In doing so, we focused on the AUC statistic not only to help facilitate the presentation and interpretation of the results, but also because it is relatively robust across different base rates and selection ratios (Smith, 1996). Using the method developed by DeLong, DeLong, and Clarke-Pearson (1988), we tested the equality of the ROC areas for concurrent and predictive validity across each of the 13 domains by gender and race/ethnicity. While gender was measured as a binary variable (1 = male; 0 = female), race/ethnicity had five categories: 1) White, 2) Black, 3) Native American, 4) Hispanic, and 5) Asian. Individuals who identified as Hispanic were placed in that category, while those identifying as Non-Hispanic were assigned to one of the four remaining categories.

To assess the substantive importance of the results, we relied on the guidelines provided by Rice and Harris (2005). Effect sizes are considered large if the value is 0.371 or higher for the correlation coefficient and 0.714 or higher for the AUC. Effect sizes are medium if the value ranges from 0.243-0.370 for the correlation coefficient and 0.639-0.713 for the AUC. Effect sizes are small if the value ranges from 0.10-0.242 for the correlation coefficient and 0.556-0.638 for the AUC. In order for a domain to be considered a criminogenic need for the MnDOC population, it must achieve at least a small, statistically significant effect size for the concurrent and predictive validity analyses. Thus, the value will need to be at least 0.100 for the correlation coefficient and 0.556 for the AUC.

Results

As shown in Table 1, which provides the descriptive statistics for our sample, 84% were male and the remaining 16% were female. Similar to the racial/ethnic breakdown for the MnDOC prison population in general, nearly half of the respondents were Non-Hispanic White individuals while the other half were Black, Indigenous, People of Color. Consistent with the design of the MnSTARR 2.0 (Duwe, 2021), a little more than 40% of the respondents were rated as either “very high” or “high” risk for recidivism. Of the 2.095 respondents who completed the assessment, there were 1,147 who had been released from prison by the end of 2022. With a follow-up period than ranged from 6-26 months, 18% of these releases had been reconvicted for a new offense by June 30, 2023.

INSERT TABLE 1 ABOUT HERE

In Table 2, we present results that analyzed the association between the 13 needs and responsivity domains and MnSTARR 2.0 risk levels and recidivism. In addition to denoting which results were statistically significant at the .01 and .05 levels, we bolded those in which

Table 1. Sample Descriptive Statistics

Predictors	*Predictor Description*	*Mean/%*	*Range*	*SD*	*N*
Substance Use Disorder	Substance Use Disorder (SUD) Need Level				
High	High Need	0.696	0-1		1,458
Medium	Medium Need	0.044	0-1		93
Low	Low Need	0.260	0-1		544
Education	Education Need Level				
High	Less than a secondary degree	0.161	0-1		337
Medium	Secondary degree or diploma	0.644	0-1		1,350
Low	Post-secondary degree or certificate	0.195	0-1		408
Anti-Social Thinking	Average overall TCU-CTS score	22.265	6-43	5.237	
Self-Identity	Self-identity scales	9.248	4-18	1.896	
Employment Scale	Employment scale	33.590	15-67	8.773	
Housing/Homelessness	Housing/homelessness scale	1.978	0-10	1.973	
Family/Domestic	Family/domestic score	5.897	2-11	1.885	
Anti-Social Peers	Attitudes towards Associates scale	35.737	2-50	10.469	
Childhood Trauma	Adverse Childhood Experiences (ACE) score	3.796	0-10	2.943	
Mental Health	Mental health score	0.543	0-7	0.858	
Religiosity	Religious Faith/Spirituality Scale	9.39	0-18	5.73	
Motivation	Motivation score	8.445	0-10	1.971	
Learning Style	Preferred learning style				
Kinesthetic	Kinesthetic learning style score	2.517	0-8	1.412	
Audio	Audio learning style score	1.997	0-7	1.394	
Read	Read learning style score	2.471	0-8	1.342	
Visual	Visual learning style score	0.947	0-6	0.947	
Gender	Men = 1; Women = 0	0.839	0-1	0.367	
Age	Age at Survey	38.397	17-81	10.717	
Race/Ethnicity	White is the reference category				
White	White	0.473	0-1	0.499	
Black	Black	0.291	0-1	0.455	
Hispanic or Latino/a	Hispanic or Latino/a	0.089	0-1	0.284	
American Indian	American Indian	0.126	0-1	0.332	
Asian/Native Hawaiian	Asian or Native Hawaiian	0.030	0-1	0.171	
MnSTARR 2.0	MnSTARR 2.0 Risk Levels				
Very High	Very High	0.216	0-1		450
High	High	0.216	0-1		450
Medium	Medium	0.190	0-1		397
Low	Low	0.381	0-1		798
Recidivism	Reconviction after release from prison	0.181	0-1	0.385	1,147
N		2,095			

Notes: SD = Standard Deviation; TCU = Texas Christian University; CTS = Criminal Thinking Scales; MnSTARR = Minnesota Screening Tool Assessing Recidivism Risk; ACE = adverse childhood experience

the value met or surpassed the small effect size threshold provided by Rice and Harris (2005). The concurrent validity results indicate that nine of the domains achieved at least a small effect size based on the correlation coefficient. Of these domains, eight also had at least a small effect size according to the AUC. Although the correlation coefficient for Education

was above 0.10, the AUC value (0.537) was below 0.556. With an AUC value of 0.640 and a correlation coefficient of 0.255, the Anti-Social Peers domain had the strongest association with MnSTARR 2.0 risk level. The remaining four domains—Mental Health, Religiosity, Motivation and Learning Style—did not achieve a small effect size with either metric.

Table 2. Concurrent and Predictive Validity for Needs and Responsivity Assessment

Domain	*MnSTARR*		*Recidivism*	
	r	AUC	r	AUC
Substance Use Disorder	**0.196****	**0.571****	0.030	0.519
Education	**0.116****	0.539	0.093**	**0.563****
Anti-Social Thinking	**0.180****	**0.579****	0.073*	**0.565****
Self-Identity	**0.141****	**0.583****	**0.130****	**0.589****
Employment	**0.112****	**0.563****	0.083**	**0.557***
Housing/Homelessness	**0.213****	**0.604****	0.**126****	**0.590****
Family/Domestic	**0.144****	**0.571****	0.078**	**0.560****
Anti-Social Peers	**0.255****	**0.640****	**0.147****	**0.619****
Childhood Trauma	**0.106****	**0.560****	0.026	0.521
Mental Health	0.058**	0.536**	-0.003	0.493
Religiosity	0.041	0.528*	0.017	0.525
Motivation	-0.050*	0.461**	-0.057	0.457
Learning Style				
Kinesthetic	0.034	0.516	-0.044	0.466
Audio	0.064**	0.517	0.078**	0.554*
Read	**-0.102****	0.463**	-0.091**	0.433*
Visual	0.019	0.509	0.069*	0.539
MnSTARR 2.0			**0.254****	**0.703****
N	2,095		1,147	

Notes: r = correlation coefficient; AUC = Area Under the Curve; MnSTARR = Minnesota Screening Tool Assessing Recidivism Risk

** $p < .01$

* $p < .05$

The predictive validity results for recidivism show that seven of the domains had an AUC value higher than 0.556, while three had a correlation coefficient above 0.10. Once again, the Anti-Social Peers domain had the largest correlation coefficient and AUC values. The MnSTARR 2.0 risk level had a AUC value of 0.703 and a correlation coefficient of 0.254. Although Substance Use and Childhood Trauma had significant, small effect size results for assessed risk level, neither one had a significant association with observed

recidivism. As with the concurrent validity findings, the Mental Health, Religiosity, Motivation and Learning Style domains did not achieve at least a small effect size with either metric.

Table 3. Concurrent and Predictive Validity by Gender

Domain	*MnSTARR*		*Recidivism*	
	AUC		AUC	
	Females	Males	Females	Males
Substance Use Disorder	0.547	**0.581**	0.472	0.522
Education	**0.586**	0.534	**0.651**	0.544*****
Anti-Social Thinking	**0.557**	**0.589**	**0.575**	**0.558**
Self-Identity	**0.588**	**0.590**	**0.673**	**0.573**
Employment	0.549	0.541	**0.613**	0.522
Housing/Homelessness	**0.630**	**0.605**	**0.615**	**0.584**
Family/Domestic	**0.595**	**0.581**	**0.619**	**0.563**
Anti-Social Peers	**0.667**	**0.636**	**0.625**	**0.621**
Childhood Trauma	**0.567**	**0.561**	0.550	0.526
Mental Health	0.524	0.551	0.495	0.515
Religiosity	0.551	0.507	**0.613**	0.498*
Motivation	0.506	0.445	0.364	0.477
Learning Style				
Kinesthetic	0.563	0.509	**0.559**	0.453
Audio	0.426	0.530**	0.466	**0.567**
Read	0.516	0.445**	0.488	0.420
Visual	0.474	0.527	0.444	**0.557**
MnSTARR 2.0			**0.732**	**0.679**
N	337	1,758	205	942

Notes: AUC = Area Under the Curve; MnSTARR = Minnesota Screening Tool Assessing Recidivism Risk

** $p < .01$

* $p < .05$

In Table 3, we compare the concurrent and predictive validity results for men and women. As with Table 2, the results were bolded if the AUC value was 0.556 or higher. Here, however, we report whether the difference in ROC curves (i.e., AUC values) for men and women is statistically significant at the .05 level. The results for concurrent validity indicate the ROC curves for men and women were significantly different for two of the Learning Style sub-domains—Audio and Read. Likewise, the predictive validity findings

Table 4. Concurrent and Predictive Validity by Race/Ethnicity

Domain	*MnSTARR*					*Recidivism*				
	White	Black	Nat. Am.	Hispanic	Asian	White	Black	Nat. Am.	Hispanic	Asian
Substance Use Disorder	**0.577**	**0.574**	0.544	0.536	**0.691**	0.524	0.462	0.512	**0.586**	**0.667**
Education	0.514	**0.563**	**0.559**	**0.561**	0.543	**0.562**	0.554	**0.557**	0.550	0.548
Anti-Social Thinking	**0.575**	**0.590**	**0.595**	**0.571**	0.529	0.530	**0.661**	0.519	0.450	0.548*
Self-Identity	**0.573**	**0.621**	**0.586**	**0.652**	**0.617**	**0.574**	**0.658**	**0.574**	0.466	**0.644**
Employment	0.544	0.522	0.547	0.540	0.519	0.508	**0.561**	**0.582**	0.369	0.518
Housing/Homelessness	**0.607**	**0.588**	**0.596**	**0.627**	0.536	**0.587**	**0.622**	0.506	**0.718**	**0.591**
Family/Domestic	**0.570**	**0.576**	**0.632**	**0.594**	**0.672**	0.553	**0.597**	0.532	**0.627**	**0.829***
Anti-Social Peers	**0.641**	**0.607**	**0.584**	**0.657**	**0.704**	**0.597**	**0.639**	**0.628**	0.531	**0.865**
Childhood Trauma	**0.561**	0.530	**0.597**	**0.606**	0.455	0.510	**0.564**	0.457	**0.695**	0.513
Mental Health	0.536	0.522	**0.637**	**0.583**	0.415**	0.498	0.520	0.481	0.504	0.532
Religiosity	0.522	0.518	0.528	**0.572**	**0.560**	0.497	0.544	0.531	0.505	**0.571**
Motivation	0.465	0.447	0.466	0.469	0.411	0.483	0.388	0.509	0.547	0.540
Learning Style										
Kinesthetic	0.514	0.530	**0.556**	**0.611**	0.430	0.468	0.432	**0.557**	0.506	0.361
Audio	0.517	0.530	**0.556**	0.461	**0.614**	0.515	**0.575**	0.554	**0.634**	**0.750**
Read	0.468	0.465	0.402	0.426	0.466	0.476	0.427	0.381	0.267	0.312
Visual	0.525	0.500	0.478	0.446	**0.591**	0.546	**0.584**	0.465	0.496	0.520
MnSTARR 2.0						**0.693**	**0.664**	**0.662**	**0.737**	**0.818**
N	1,048	662	197	118	57	589	340	130	61	27

Notes: AUC = Area Under the Curve; MnSTARR = Minnesota Screening Tool Assessing Recidivism Risk

** $p < .01$

* $p < .05$

indicate the ROC curves were significantly different for men and women for the Education and Religiosity domains. More specifically, the scores for these two domains were more predictive of recidivism for women than for men; that is, lower levels of educational achievement and religiosity were more strongly associated with recidivism for the women.

Similar to Table 3, the results in Table 4 not only provide the AUC values by race/ethnicity across the 13 domains for concurrent and predictive validity, but also whether any statistically significant differences in ROC curves were observed. The concurrent validity results indicate that statistically significant differences were found for one domain—Mental Health. In particular, whereas Mental Health was positively associated with recidivism risk level for Native American individuals, it was negatively associated for people identifying as Asian.

For predictive validity, the results indicate that ROC curves were significantly different for the Anti-Social Thinking and Family/Domestic domains. More specifically, Anti-Social Thinking had a positive, medium effect size for Black individuals, while it was negatively associated with recidivism for Hispanic people. Likewise, the Family/Domestic domain was not as predictive of recidivism for Hispanic individuals as it was for everyone else, especially those identifying as Asian.

Discussion

The results were, for the most part, consistent with what we hypothesized. All seven of the domains we identified as criminogenic needs had at least one significant, positive association for the concurrent and predictive validity analyses. Among these criminogenic needs, Anti-Social Peers had the strongest association with both the MnSTARR 2.0 risk level and actual recidivism. Of the six domains we initially identified as specific responsivity factors,

the Mental Health, Religiosity, Motivation, and Learning Style domains did not have significant, positive associations with assessed and observed recidivism, which is consistent with what we hypothesized.

Contrary to our hypotheses, however, Self-Identity not only had significant, positive results in both the concurrent and predictive validity analyses, but it also had one of the stronger associations with assessed and observed recidivism. Our findings suggest that, at least for Minnesota's prison population, Self-Identity is a criminogenic need rather than a specific responsivity factor. Over the last few decades, a growing body of literature has shown that identity transformation is important to the desistance process (Paternoster & Bushway, 2009; Rocque et al., 2016). The results presented here suggest that people who retain their anti-social identities are more likely to reoffend and be assessed as high risk, while those who have shed this identity in favor of one that is more pro-social (i.e., replacement self) are less likely to recidivate. Because identities are dynamic (i.e., they can change over time) and have a significant, direct impact on recidivism, our findings suggest that Self-Identity should receive greater consideration as a distinct criminogenic need.

While we hypothesized that Childhood Trauma is a specific responsivity factor, the concurrent and predictive validity results were mixed. This domain had a small effect size, albeit barely, for assessed recidivism risk, whereas it did not have a significant association with observed recidivism. Although statistically significant differences were not found for Childhood Trauma across gender and racial/ethnic categories, both the concurrent and predictive validity results suggest that it may be more associated with recidivism for Hispanic individuals.

Along the same lines, the findings indicated that education and religiosity were more predictive of recidivism for the women than for the men. While Anti-Social Thinking was predictive of recidivism for Black individuals, the Family/Domestic domain was not strongly associated with reoffending for White and Native American people. On the whole, however, gender and racial/ethnic differences in the concurrent and predictive validity were relatively minimal across the 13 domains.

Conclusion

The findings from this study are promising, although there are several limitations worth noting. First, at a little under 18 months, the average follow-up period for measuring recidivism was relatively brief. Second, although we attempted to evaluate racial/ethnic differences in the assessment system's performance, the sample sizes for some of the categories (e.g., Hispanic and Asian) were small, reducing the confidence that can be placed in those findings. Finally, because our sample consisted of individuals from one state's prison system, the extent to which the results are generalizable to other correctional populations is unclear.

Despite these limitations, the findings from this study have several notable implications for policy, practice and future research. First, the results suggest the MnDOC has a needs assessment that, for the most part, performs as expected. Thus, when the MnDOC implements the self-reported assessment, it will have a validated needs assessment system for its population. When paired with the MnSTARR 2.0, which has also been validated on Minnesota's prison population, the results from these RNR assessments should lead to the development of more effective case plans, which will, in turn, enhance the delivery of services, programs, and interventions for people imprisoned in the MnDOC.

Second, the results from this study highlight the potential for using self-reported assessment processes. Widely-used assessments, including the Level of Service tools previously used by the MnDOC, require staff to manually score the instrument through a labor-intensive, face-to-face interview. While correctional agencies in the U.S. have long been understaffed, the onset of the COVID-19 pandemic further exacerbated staffing shortages. For agencies still struggling to achieve adequate staffing levels, the use of self-reported assessments provides a more efficient process that requires minimal staff resources. Moreover, given that the MnDOC uses an automated scoring process for its risk assessment instrument, it is anticipated that virtually everyone in Minnesota's prison system will be assessed for risk, needs and responsivity.

Third, while this study has provided evidence on the concurrent and predictive validity of the 13 domains, future research is needed on the association between domain scores and completion of programming. In particular, are the domains, especially those for specific responsivity, predictive of programming participation and completion? Does the association vary by the type of program? Along the same lines, what services, programs and interventions have an impact on the criminogenic need domains? Empirically addressing questions such as these would presumably lead to the development of better case plans, program delivery and, ultimately, recidivism outcomes.

Finally, more validation studies are needed for instruments that assess for specific responsivity factors. To be sure, assessment tools used for correctional populations do not always assess for responsivity, and those that do often consist of little more than a checklist of items such as gender, language, or culture. Nevertheless, we suggest there should be greater clarity and rigor in determining whether an instrument, or domain within an

instrument, achieves validity in assessing responsivity. Unlike a criminogenic need, a specific responsivity factor is not expected to have a direct, significant relationship with recidivism. It is anticipated, however, that a specific responsivity factor will either impede or facilitate involvement and completion of programming. To this end, future research should devote more attention to examining whether specific responsivity factors have an empirical relationship with program involvement and completion.

References

Andrews, D. A., Bonta, J., & Wormith, S. J. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency*, *52*(1), 7–27. https://doi.org/10.1177%2F0011128705281756

Bonta, J., & Andrews, D. A. (2017). *The psychology of criminal conduct*. Routledge.

Caudy, M.S., Durso, J.M., & Taxman, F.S. (2013). How well do dynamic needs predict recidivism? Implications for risk assessment and risk reduction. *Journal of Criminal Justice, 41*, 458-466.

Clark, V.A. (2016). Predicting two types of recidivism among newly released prisoners: First addresses as "launch pads" for recidivism or reentry success. *Crime & Delinquency*, 62(10), 1364–1400.

Clark, V.A. & Duwe, G. (2023). Evaluating the convergent and discriminate validity of a needs and responsivity assessment system [Manuscript under review]. Minnesota Department of Corrections.

Cullen, F. T. (2002) "Rehabilitation and Treatment Programs." In J. Q. Wilson and J. Petersilia (eds.), *Crime: Public Policies for Crime Control*, 2nd edition. San Francisco: ICS Press.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receive operating characteristic curves: A nonparametric approach. *Biometrics*, *44*, 837-845.

Duwe, G. (2014). The development, validity, and reliability of the Minnesota Screening Tool Assessing Recidivism Risk (MnSTARR). *Criminal Justice Policy Review, 25*, 579-613.

Duwe, G. (2021). Evaluating bias, shrinkage, and the home-field advantage: Results from a revalidation of the MnSTARR 2.0. *Corrections: Policy, Practice and Research*. Advance online publication. https://doi.org/10.1080/23774657.2021.2011802

Duwe, G. & Rocque, M. (2017). The effects of automating recidivism risk assessment on reliability, predictive validity, and return on investment (ROI). *Criminology & Public Policy, 16*, 235-269.

Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., & Marks, J. S. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: The Adverse Childhood Experiences (ACE) Study. *American Journal of Preventive Medicine, 14(4)*, 245-258.

Fleming, N., & Mills, C. (1992). Not another inventory, rather a catalyst for reflection. *To Improve the Academy*, *11*, 137-155.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology, 34*, 575-607.

Gendreau, P., French, S.A. & Gionet, A. (2004). "What Works (What Doesn't Work): The Principles of Effective Correctional Treatment." *Journal of Community Corrections, 13 (*4-6), 27-30.

Giordano, P. C., Cernkovich, S. A., & Rudolph, J. L. (2002). Gender, crime, and desistance: Toward a theory of cognitive transformation. *American Journal of Sociology*, 107(4), 990-1064.

Kirk D.S., Barnes G.C., Hyatt J.M., & Kearley, B.W. (2018). The impact of residential change and housing stability on recidivism: Pilot results from the Maryland Opportunities through Vouchers Experiment. *Journal of Experimental Criminology*, *14(2)*, 213–226.

Jang, S.J., Johnson, B.R., Hays, J., Hallett, M., & Duwe, G. (2017). Religion and Misconduct in "Angola" Prison: Conversion, Congregational Participation, Religiosity, and Self-Identities, *Justice Quarterly*, DOI:10.1080/07418825.2017.1309057.

Johnson, B. R., Jang, S. J., Larson, D. B., and De Li, S. (2001). Does adolescent religious commitment matter? A reexamination of the effects of religiosity on delinquency. *Journal of Research in Crime and Delinquency*, 38: 22-44.

Knight, K., Garner, B.R., Simpson, D.D., Morey, J.T., & Flynn, P.M. (2006). An assessment for criminal thinking. *Crime & Delinquency*, *52*(1), 159-177. https://doi.org/10.1177%2F0011128705281749

Leite, W. L., Svinicki, M., & Shi, Y. (2010). Attempted validation of the scores of the VARK: Learning styles inventory with multitrait–multimethod confirmatory

factor analysis models. *Educational and psychological measurement*, *70*(2), 323-339.

McCormick, S., Peterson-Badali, M., & Skilling, T. A. (2017). The role of mental health and specific responsivity in juvenile justice rehabilitation. *Law and Human Behavior*, *41*, 55–67.

McIntire, S. A., & Miller, L. A. (2005). *Foundations of psychological testing (2nd ed.)*. Thousand Oaks, CA: Sage Publishing.

Mei, X., Hamilton, Z., Kowalski, M., & Kigerl, A. (2021). Redesigning the Central Eight: Introducing the M-PACT Six. *Youth violence and juvenile justice*, *19(4)*, 445-470.

Mills, J. F., Kroner, D. G., & Forth, A. E. (2002). Measures of Criminal Attitudes and Associates (MCAA): Development, factor structure, reliability and validity. *Assessment*, *9*, 240-253.

Mitchell, O., Wilson, D.B., & MacKenzie, D.L. (2007). Does incarceration-based drug treatment reduce recidivism? A meta-analytic synthesis of the literature. *The Journal of Experimental Criminology*, *3*, 353-375.

Mowen, T.J., Stansfield, R., & Boman, J.H. (2018). During, after or both? Isolating the effect of religious support on recidivism during reentry. *Journal of Quantitative Criminology*, *34*, 1,079-1,101.

Paternoster, R., & Bushway, S. (2009). Desistance and the" feared self": Toward an identity theory of criminal desistance. *The Journal of Criminal Law and Criminology*, 1103-1156.

Pinals, D.A., Gaba, A., Shaffer, P.M., Andre, M.A., & Smelson, D.A. (2021). Risk-need-responsivity and its application in behavioral health settings: A feasibility study of a treatment planning support tool. *Behavioral Science and the Law*, *39*, 44-64.

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law and human behavior*, *29*(5), 615-620.

Rocque, M., Posick, C., & Paternoster, R. (2016). Identities through time: An exploration of identity change as a cause of desistance. *Justice Quarterly*, *33*(1), 45-72.

Sachs, N.M. & Miller, J. (2018). Beyond responsivity: Client service engagement in a

reentry demonstration program. *International Journal of Offender Therapy and Comparative Criminology*, *62(13)*, 4295-4313.

Sease, T. B., & Knight, K. (2022). Development and Testing of the Texas Christian University Criminal Thinking Scales 3.0. *Crime & Delinquency*, DOI: 10.1177/00111287221134917.

Smith, W. (1996). The effects of base rate and cutoff point choice on commonly used measures of association and accuracy in recidivism research. *Journal of Quantitative Criminology*, *12*, 83-111.

Sperber, K.G., Latessa, E.J., & Makarios, M.D. (2013). Examining the Interaction between Level of Risk and Dosage of Treatment. *Criminal Justice and Behavior, 40*, 338-348.

Steiner B., Makarios M.D., & Travis L. (2015). Examining the effects of residential situations and residential mobility on offender recidivism. *Crime & Delinquency, 61(3),* 375–401.

Taxman, F. S., Giuranna, R. A., & Dumenci, L. (2011). Construct and predictive validity of criminal thinking scales. *Criminal Justice and Behavior*, *38*(2), 174-187. https://doi.org/10.1177%2F0093854810389550.

Zimet, G. D., Dahlem, N. W., Zimet, S. G., & Farley, G. K. (1988). The multidimensional scale of perceived social support. *Journal of personality assessment*, *52*(1), 30-41.