The Massachusetts Youth Screening Instrument-Version 2 (MAYSI-2):

Comprehensive Research Review

Thomas Grisso
University of Massachusetts
Medical School

Samantha Fusco
Fordham University

Melissa Paiva-Salisbury
University of Vermont

Rachael Perrault
University of Massachusetts
Medical School

Valerie Williams
University of Massachusetts
Medical School

Richard Barnum
Private Practice, Boston, MA

This review describes the results of all published research reports and dissertations through September 2011 that have evaluated or used the *Massachusetts Youth Screening Instrument-Version 2* (MAYSI-2). The instrument was released in 2000 and is now widely used in the U.S. for brief mental health screening at entry to juvenile justice programs. The review describes the results of studies using the MAYSI-2 to examine demographic groups, the psychometric properties of the tool, correlations with similar measures, and the relation of MAYSI-2 scale scores to behavioral, clinical and offense characteristics of young people admitted to juvenile justice programs. Results are used to offer cautions for clinical and research use of the instrument, as well as suggestions for future research that could improve clinical use of the MAYSI-2 in mental health screening in juvenile justice settings.

Cite as: Grisso, T., Fusco, S., Paiva-Salisbury, M., Perrault, R, Williams, V., & Barnum, R. (2012). *The Massachusetts Youth Screening Instrument-Version 2 (MAYSI-2): Comprehensive research review*. Worcester, MA: University of Massachusetts Medical School. www.nysap.us

The Massachusetts Youth Screening Instrument-Version 2 (MAYSI-2):

Comprehensive Research Review

The *Massachusetts Youth Screening Instrument-Version 2* (MAYSI-2) (Grisso &
Barnum, 2006; Grisso et al., 2001) is a brief mental health screening tool for routine
administration at entry to any juvenile justice facility or service. Its purpose is to identify youth
who may need immediate attention regarding possible suicide risk and emergent mental health
and substance use needs.   Released 12 years ago (Grisso & Barnum, 2000), the MAYSI-2 is
now registered for use in over 2,000 sites in 47 states, including statewide use in all intake
probation, detention and/or corrections facilities in 44 states.   Researchers have translated the
MAYSI-2 into 13 languages. During this period of rapid adoption of MAYSI-2, substantial
research has examined the structure, utility, and validity of the instrument.

This is the first comprehensive review of the body of research studies that have examined
the properties of the MAYSI-2 and have used it to describe delinquent youth and test hypotheses
about their behavior, clinical conditions, and offending.   The review has two objectives.   For
researchers, it identifies psychometric characteristics of the tool, the range of approaches to
MAYSI-2 validation in existing studies, and research questions not yet addressed.   For
practitioners, it provides evidence supporting and limiting the use of the MAYSI-2 when used in
juvenile justice settings, especially interpretations to avoid.

The MAYSI-2 is fully described in its manual (Grisso & Barnum, 2006) and summarized
in Grisso and Quinlan (2005). To assist readers who are not familiar with the MAYSI-2, a basic
summary in the Appendix to this article provides a description of its history, nature and
development. It was designed specifically as a self-report 15-minute screening (triage) tool to be

administered, often by non-clinical personnel, to all youth at the time of intake (within 1-3 hours after admission) in juvenile probation offices, juvenile pre-trial detention centers, and juvenile justice corrections and residential facilities. Its purpose is to identify symptoms (represented by thoughts, feelings and behaviors) that are found in many psychiatric diagnostic conditions of youth, but it was not intended to provide diagnoses. Ideally it offers information to intake staff that identifies young people who might require an immediate mental health response (e.g., suicide precautions, need for further evaluation, referral for clinical consultation), but it was not developed to provide long-term treatment recommendations.

Table 1 offers a description of the 7 MAYSI-2 scales and the acronyms that are used to identify them throughout this review. Readers should be aware that there are two ways to express MAYSI-2 scores for research and clinical purposes. Some research has used continuous scores and means on each scale, while others have used the standardized cut points on MAYSI-2 scales to express low/high scores dichotomously. (Table 1 provides brief descriptions of those cut points, and the Appendix describes their derivation and logic.) There is no "total MAYSI-2 score" in the MAYSI-2 system.

———————————

Insert Table 1 about here

———————————

The review has 5 sections examining empirical research findings: (a) description and comparison of demographic groups on MAYSI-2 scales, (b) studies of MAYSI-2 administration and response styles, (c) psychometric properties of the MAYSI-2 scales, (d) comparisons to other measures of similar constructs, and (e) studies providing inferences about the tool's

predictive and construct validity.   The Discussion suggests future areas of MAYSI-2 research and methodological cautions, as well as guidance for juvenile justice personnel when employing mental health screening of young people with the MAYSI-2.

Reports in this review were retrieved with a thorough search of electronic databases of scientific journal publications and doctoral dissertations that employed the MAYSI-2 and were published prior to September 1, 2011.   We identified 56 reports meeting these criteria. Excluding the three largest samples (70,423 in Vincent et al., 2008; 54,716 in Maney, 2011; 18,607 in Cauffman, 2004), sample sizes ranged from 29 to 7,746 (average = 1,336), with the 25th to 75th percentiles for the distribution of sample sizes being 228 to 1,840.   Collectively the studies represented over 220,000 MAYSI-2 administrations.

The review makes frequent reference to the "MAYSI-2 National Norms" as a point of comparison.   Data for the National Norms (reported in Grisso & Barnum, 2006: Manual Supplement) included over 70,000 cases contributed by 283 juvenile justice facilities in 19 states distributed nationwide, representing consecutive admissions to those facilities during 2003-2005.

Demographic Descriptions and Comparisons

This section first describes the demographic composition of the studies collectively, then reviews MAYSI-2 comparisons by gender, age, race/ethnicity, and juvenile justice setting.

General Demographic Composition

Almost all studies reported gender of cases, and collectively they included 79% boys and 21% girls.   The racial/ethnic composition collectively was 33% black, 31% white, 18% Hispanic/Latino, 3% Other (primarily Native Hawaiian, Native American, Pacific Islander), 1% Asian, .01% Bi-Racial, and 14% unreported by race/ethnicity.   The great majority of studies employed samples obtained in a single type of juvenile justice setting.   Pre-trial detention

samples and juvenile corrections samples were the most frequent and about equally represented, with fewer (about 10% each) reporting on intake probation samples or on community or residential non-juvenile-justice samples. Less than 10% of the studies used samples that included more than one type of juvenile justice setting.

Gender Effects

In the MAYSI-2 National Norms, significantly greater proportions of girls than boys scored above Caution and Warning cutoffs on every scale except ADU (Alcohol/Drug Use). Twelve studies in this review offered between-gender comparisons of MAYSI-2 scores, some using continuous scores and others Caution cut-off dichotomous classification. All reported on the main MAYSI-2 clinical scales: ADU, AI, DA, SC and SI. TD (Thought Disturbance) usually was not reported in gender comparisons because there is no TD scale for girls (see Appendix for explanation), and TE was less often used (because it has no cut-off scores in the MAYSI-2 system).

Of the 12 studies, significantly higher scores for girls were found on all 5 of the above scales by Cauffman (2004; also see Archer et al., 2004, which did not report statistical tests of significance) and on all but ADU in 5 other studies (Cauffman et al., 2007; Grisso, et al., 2001; Hayes et al., 2005; Kerig et al., 2009; Maney, 2011). Three studies found significantly higher means for girls on 3 scales—DA, SC, and SI (Nordness et al., 2002; Stathis et al., 2008), and ADU, SC, and AI (Archer et al., 2010)—and 1 study on DA and SI only (Gretton & Clift, 2011).

Vincent et al. (2008) performed a meta-analysis of the National Norms dataset of 283 sites to examine the generalizability of gender differences across sites. Among the 5 scales in question, girls were significantly higher than boys on all scales except ADU (on which girls were slightly lower). Gender effect sizes were moderate to large, and the result was sufficiently homogeneous to be relevant for general application nationwide. One ADU gender x age effect

was reported; among younger adolescents, girls were higher than boys on ADU (with a small effect size), while older girls and boys did not differ in proportions over ADU Caution cut-off.

Age Effects

Only 4 reports presented MAYSI-2 results by age (Archer at al., 2004; Cauffman, 2004; Grisso, et al., 2001; Vincent et al., 2008), all reporting significantly lower means on ADU for younger than for older adolescents. The only studies to find other significant gender differences were Cauffman (2004), with younger adolescents obtaining a significantly greater mean AI, and Grisso et al. (2001) who found a significant difference on TE (with older adolescents higher).

Race/Ethnicity Effects

Race/ethnicity comparisons on MAYSI-2 scales were reported in 10 studies in the present review (Archer et al., 2004; Cauffman, 2004; Cauffman & MacIntosh, 2006; Dalton et al., 2009; Grisso et al., 2001; Maney, 2011; Osterlind et al., 2007; Stathis et al., 2008; Stewart & Trupin, 2003; Vincent et al., 2008).   All of these studies compared white and black samples (except Stathis et al., 2008, who reported on indigenous and non-indigenous Australian youth), 6 reported results for Hispanic youth (Cauffman, 2004; Cauffman & MacIntosh, 2006; Grisso et al., 2001; Maney, 2011; Stewart & Trupin, 2003; Vincent et al., 2008), and 3 compared Asian results (Cauffman & MacIntosh, 2006; Maney, 2011; Vincent et al, 2008).

The collective results of these studies are complex so they will be reported here primarily for comparisons of white and black youth. Across studies, white youth had somewhat higher means or Caution cut-off percents on all scales than did black youth (in the Stathis et al. study in Australia, non-indigenous youth higher than indigenous youth).   But among studies offering tests of significance (2 did not), none found significant differences between white and black youth on DA and TD.   Results varied across studies, however, for the other scales. The most

consistent finding (6 out of 8 studies) was a significantly higher mean (or percentage above Caution cut-off) on ADU for white compared to black youth. Higher scores for white than for black youth on SC was found in 4 studies, on SI in 2 studies, and on AI in 2 studies.

A substantial lack of consistency in white and black differences across sites was reported by Vincent et al. (2008) in their meta-analysis of the National Norm data from 283 sites nationwide. They found small to significant effect sizes (white higher than black) for 3 scales, most notably a moderately-strong effect size for ADU. Yet their test of homogeneity of this effect across sites indicated that the effect could not be generalized. Higher ADU scores for white than for black youth were found consistently, but the magnitude of the differences varied widely across sites, ranging from substantial to non-significant. Results for Hispanic youth tended to fall between those of white and black youth.

Four studies reported MAYSI-2 results for race/ethnicity by gender (Archer et al, 2004; Cauffman, 2004; Grisso et al., 2001; Maney, 2011). All reported that white, black and Hispanic girls scored higher than boys of their own race/ethnicity on all scales except ADU. In Cauffman's (2004) study, white girls scored higher than black or Hispanic girls on all MAYSI-2 scales with the exception of DA (higher for Hispanic girls), and black girls scored lower on all scales than the other two race/ethnic girl samples.

Juvenile Justice Setting

The original MAYSI-2 norms (Grisso et al., 2001) reported significantly lower ADU scores for intake probation youth than for detention and corrections youth, as well as significantly lower SI mean scores for corrections youth than intake probation and detention youth. Maney (2011) found few substantial differences between intake probation, detention and

corrections settings, with the exception of significantly higher TE scores for youth in juvenile corrections settings.

Archer et al. (2010) found no differences between detention and corrections samples on any MAYSI-2 scale. Comparing youth in juvenile detention to youth in the community (not currently involved in juvenile justice), Cauffman et al. (2007) found significantly higher scores for detention youth on all MAYSI-2 scales. Murrie et al. (2009) compared youth in juvenile corrections to youth who were in adult prison, finding significantly higher means for the juvenile corrections group on all MAYSI-2 scales.

Summary

Consistency across studies was apparent regarding higher MAYSI-2 scores generally for girls than for boys, lower ADU scores for younger adolescents, and generally lower scores for black youth in juvenile justice settings than for white youth. Nevertheless, considerable variability appeared across studies in the magnitude of differences between different race/ethnic groups, and regarding the specific scales on which the differences were significant. This variability probably arises because most of the studies were performed in single sites or states. The 283-site meta-analytic study by Vincent et al. (2008) verified considerable heterogeneity among sites, especially for race/ethnicity-related patterns of MAYSI-2 scores. Theoretically, there are many reasons why juvenile justice sites might vary in circumstances that could produce inter-site variability in mean scores on behavioral and mental health measures. These are described in the Discussion section. In a few cases, however, comparisons simply were misinterpreted. For example, one study reported a sample's substantial mean difference from the National Norms on ADU (McConihay, 2008), but had defined the "Caution range" as youth scoring between the Caution and Warning cut-off, while the standard use of the Caution cut point

is to identify the percentage above the Caution cut score (which always includes cases above the Warning cut score as well).

Administration and Response Style

As with any self-report instrument, MAYSI-2 responses might differ based on administration variations across sites: for example, time between facility admission to time of testing, differences in instructions provided (e.g., about future uses of the information), and whether administered in person or by computer. Hayes et al. (2005) found little difference in scores for MAYSI-2 results obtained with the computerized version of MAYSI-2 (called MAYSIWARE) and the paper-and-pencil version. With that exception, no other studies have examined administration variables and MAYSI-2 performance.

Two studies raised questions about response styles when youth complete the MAYSI-2. Maney (2011) found that about 7-8% of her sample of over 54,000 MAYSIWARE administrations were completed in 2 minutes or less, which is fairly difficult to do if one is reading each item. Moreover, a small but disturbing percentage of youths provided "no" answers to all items or "yes" answers to all items. Finally, she found that about 90% of the cases with an "all no" response set were completed in less than 2 minutes. Together these facts suggest concern about the validity of MAYSI-2's completed in less than 2 minutes.

The "all no" response style was also noted by Logan (2007), who examined samples administered the MAYSI-2 during 2003-2005 in two of the same juvenile facilities that had participated in the original MAYSI validation studies in 1997-1998. She found "all no" cases in only 4% of the 1990s sample compared to 25% in the mid-2000s sample. She speculated that mental health screening was "novel" to youth in the 1990s, but in subsequent years had become routine (in that jurisdiction), so that many youth might have taken the MAYSI-2 repeatedly

within a short time span. Thus, youth might have been overexposed to the practice and simply

circled "no" to all questions in order to quickly finish taking the instrument.

<div align="center">Psychometric Properties</div>

Numerous studies and reviews examined the reliability, internal consistency and structure

of the MAYSI-2 since its original validation. All of these studies were performed with boys

and/or girls in juvenile justice settings.

Alpha Coefficients and Scale Inter-Correlations

Alpha coefficients have been reported for both paper-and-pencil and computer-assisted

MAYSI-2 formats, for overall sample, by race, and by gender (Archer et al., 2004; Grisso et al.,

2001; Ford et al., 2007; Hayes et al., 2005). Individual scales' alpha coefficients ranged from a

low of .54 to a high of .90, with most scales being above .70.   The TD scale was typically the

weakest in this regard.   Alpha coefficients have been found to be the same or only negligibly

different for white, black, and Hispanic youth, and for boys and girls, for all MAYSI-2 scales

(Archer et al., 2004; Grisso et al., 2001; Hayes et al., 2005), with the exception of TE in the

Hayes et al. study (somewhat lower for boys than girls).

Several studies reported correlations among MAYSI-2 scales (Archer et al., 2004, 2010;

Butler et al., 2007; Grisso et al., 2001; Johnson, 2006; Osterlind et al., 2007; Padilla, 2006; Vance,

2006).   Inter-correlations ranged from a low of .00 to a high of .70 for various pairs of scales.

However, considerable differences were reported across studies.   For example, out of 21 possible

pairs of MAYSI-2 scales, inter-correlations above .40 were reported in about 62% of the pairs in

one study (Vance, 2005), 50% in two others (Archer et al., 2004, 2010; Grisso et al., 2001), and

25% in another study (Butler et al., 2007).   This might be related to differences in samples; the

first was in a detention center, the second and third combined probation, detention and corrections

settings, and the fourth was a residential treatment program. Typically the scales most highly correlated with each other were DA and SI, as well as DA and AI. The latter relation between depressive symptoms and anger is consistent with diagnostic criteria and research on adolescent depression (American Psychiatric Association, 2000: Sheeber et al., 2009).

Test-Retest Reliability

Archer et al. (2004) retested youth with the MAYSI-2 after an average of 15 days, finding test-retest correlations in the moderate to high range (r = .60 to .82). Grisso and Barnum (2006) reported 6-12 day retest results similar to those of Archer et al.; the changes were not significant for girls on any scale, but were significant for boys on DA, SC and TD. Cauffman's (2004) test-retest comparison involved a much longer delay, averaging 111 days. Thus she found significantly lower reliability over time (correlations ranging from .27 to .70 across scales). Interpretation requires recognition that the MAYSI-2 was designed to identify emotions and thoughts "during the past few months" as recalled at a given time, not to measure traits or other enduring conditions.

Conceptual Structure

The current MAYSI-2 scales were determined using factor analysis of the instrument's 52 items (Grisso et al., 2001; Grisso & Barnum, 2006). Using the same methods with a different sample, Archer et al. (2004) found 7 factors that clearly replicated the MAYSI-2 scales, with congruence coefficients ranging from .80 (TE and TD) to .94 (ADU). As in the original factor analyses, they found no TD factor for girls and an eighth factor that could not be interpreted. Ford et al. (2007) found a six-factor solution, replicating all but the DA scale.

Separate factor analyses for race/ethnic groups have not been conducted. The similarity across race/ethnic groups in alpha coefficients (described earlier) suggest that separate factor

analyses might produce results similar to the above studies. Yet this would not verify that the scales operate similarly well for all race/ethnic groups. For example, Cauffman and MacIntosh (2006) and McCoy (2010, 2011) used Rasch analyses to identify "differential item functioning" (DIF), an analytic method that examines whether individual items in a scale are operating similarly (appear to be measuring the same dimension) with different classes of respondents. They found differences between black and white youth in the apparent latent meaning of 13 of the 52 items (many of them in TD and AI). McCoy explored these differences with qualitative methods, identifying several "themes:" underreporting, differences in willingness to discuss certain feelings, different interpretations of what the items were asking, and difficulty remembering the requested information.

Summary

Many psychometric characteristics described in MAYSI-2's original development have been replicated, and generally they suggest considerable internal consistency, but with some weaknesses in TD.   More will be said in the Discussion section, however, about the need for future studies to identify potential differences in meanings of some items across race/ethnic groups.

Comparison to Conceptually-Similar Measures

Several studies have examined concurrence of the MAYSI-2 scales with measures of similar mental and behavioral health constructs.   In most studies the two instruments appeared to have been administered at the same time or within a few days.

Alcohol/Drug Use (ADU)

Comparing the ADU Caution cut-point to the *Millon Adolescent Clinical Inventory* (MACI) Substance Abuse-Proneness scale, Butler et al. (2007) found a correlation of .66 (boys),

and Grisso et al. (2001) obtained an AUC of .82 (boys) and.79 (girls).   Archer et al. (2010)

found ADU correlated substantially with the *Substance Abuse Subtle Screening Inventory*.

Angry-Irritable (AI)

Grisso et al. (2001) reported strong relations between AI and the *Youth Self Report* (YSR)

Externalizing scale (r = .54) as well as the YSR Aggressive Behavior scale (r = .48).   Similar

results were found for AI and MACI Impulsive Propensity (r = .41: Butler et al., 2007; r = 45:

Grisso et al., 2001). Grisso et al. cautioned that these relations were stronger for boys than girls,

but Caldwell et al. (2006) found equally high correlations (between .50 and .70) for boys and

girls (both black and white) between AI scores and a standardized measure of impulsivity.

Depressed-Anxious (DA)

In the Grisso et al. (2001) study comparing MAYSI-2 to MACI and YSR scales, DA

scores were most closely related to the MACI's Depressed Affect scale and Suicidal Tendency

scale, as well as the YSR's Anxious-Depressed scale and Self-Destruction scale.   Butler et al.

(2007) also noted that DA, MACI Depressive Affect, and MACI Suicidal Tendency produced a

triad of correlations (.42 to .57) that were higher than their relations with any other MACI scales.

Somatic Complaints (SC)

Grisso et al. (2001) found modest correlations between SC and the YSR Somatic

Complaints scale for both boys (r = .40) and girls (r = .51).   However, SC was just as strongly

related to the YSR's Internalizing dimension (.47 for boys, .50 for girls).

Suicide Ideation (SI)

Good correlations have been reported between MAYSI-2 Suicide Ideation and scores on

the *Suicide Ideation Questionnaire* (r = .58: Chapman & Ford, 2008), the MACI Suicide

Tendency scale (Butler et al., 2007: r = .57; Grisso et al., 2001: r = .61, and AUCs of .94 for boys

and .91 for girls when using SI Caution cut-points), the YSR Self Destructive scale (r = .44: Grisso et al., 2001), the *Positive and Negative Suicide Ideation Inventory's* Negative Suicide Ideation index (girls, r = .68: Tille & Rose, 2007), and the *Children's Depressive Inventory* (girls, r = .51: Tille & Rose, 2007).   (Other studies examining the relation of SI to diagnostic criteria using suicide history and current suicidal thoughts are described in the section on construct validity.)

Thought Disturbance (TD)

The only study to examine TD in relation to other scales (Grisso et al., 2001) found a moderate correlation with the YSR Thought Problems scale (r = .40), and an AUC of .73 using the TD to predict YSR Thought Problem scores in the clinical range.

Traumatic Experiences (TE)

Chapman and Ford (2008) found a significant relation between TE scores and the *Suicide Ideation Questionnaire* (SIQ), as well as substantial correlations between certain items on the TE and similar items on the SIQ focused on similar constructs associated with trauma symptoms (e.g., intrusive re-experiencing of traumatizing event, increased sensitivity to threatening social cues).

Summary

In general, studies have found substantial relations between MAYSI-2 scales and validated tools measuring similar constructs.   Considerably more research of this type is needed on some scales, especially Thought Disturbance.

<div align="center">Studies Addressing Construct Validity</div>

Numerous studies have used MAYSI-2 scales to examine their relation to other criteria with which the scales might be expected to have theoretical or conceptual relations.   The studies are reviewed according to classes of external criteria.

Relation to Delinquent Behavior

   Recidivism.   Two studies examined repeat offending *prospectively*, Bisbee (2010) for 1 year and Harshbarger (2007) for 2 years.   In Bisbee's study, recidivism was significantly related to ADU and AI for boys and girls (as well as DA for boys and SC for girls).   In Harshbarger's survival analysis, the MAYSI-2 scores significantly improved identification of recidivism compared to use of the historical variables alone.

   Two additional studies examined MAYSI-2 scales in relation to repeat offending determined *retrospectively*: that is, current MAYSI-2 scale scores compared to past delinquency records.   In a study by Marczyk et al. (2003), entering *Psychopathy Checklist:Youth Version* scores, *Youth Level of Services* scores and a "MAYSI total score" into a logistics regression accounted for only 1% of the variance in "total violent charges" in the past.   Moreover, no single MAYSI-2 scale was related to recidivism.   However, a regression analysis entering all MAYSI-2 scales individually, together with the Prior and Current Offenses/Dispositions scale of the YLS, produced a significant model that accounted for 29% of the variance in violent charges.   In the other retrospective study, examining girls in a juvenile correctional setting, Tille and Rose (2007) found that all MAYSI-2 scales were significantly higher for girls with second convictions compared to those with first-time convictions.

   In-custody infractions.   Three studies examined the relation of MAYSI-2 scales to institutional infractions—that is, youths' behaviors that endangered others while they were in custody in juvenile facilities.   Butler et al. (2007) and Vance (2006) both reported significantly higher AI scores for youth with infractions (and in Vance, TD as well).   DeLisi et al. (2010b) reported significantly higher AI scores for youth with any recorded assault on staff.   None of

these studies translated their results into decision rules that might aid juvenile justice staff in identifying youth with elevated likelihood of in-custody infractions.

Specific violent groups. Studies have examined MAYSI-2 scores in relation to several types of violent behavior or violence potential among adolescents. Comparing youths with past offenses classified as violent or non-violent, Maney (2011) reported significantly higher AI, DA and SI scores for youth with violent offenses (both boys and girls). Further dividing the violent offenses into domestic (against family members) and non-domestic violence, the study found AI, SI and DA higher for domestic violence boy offenders, but all MAYSI-2 scales were significantly higher for the domestic violence girl offenders.

Assessing incarcerated boys for past "bullying," (frequency of threatening other youths with force), Bender et al. (2010) reported significantly higher scores on ADU and TE for youths with a bullying history. Interestingly, the TE scale also contributed significantly to a logistic regression aimed at identifying youth who had or had not been carrying weapons during their recent offense (Vaughn et al., 2006); lifetime substance use and gang fighting were the most significant factors, with TE and other evidence of past trauma having secondary significance.

Garland and Howard (2010a, 2010b) found that youths whose substance use included inhalation of computer duster spray were higher than other substance users on ADU, SI and TE. Delisi et al. (2010a, 2010b) developed several statistical models to identify sexual misconduct youth within a broader sample of offenders, finding AI (higher for sexual misconduct youth) contributing significantly to some of the models. Roe-Sepowitz (2007) and Hickle and Roe-Sepowitz (2010) reported using the MAYSI-2 to study adolescent girls charged with homicide (dividing them into "conflict event" and "crime event" homicides), and to examine adolescent

girls charged with arson (alone, or with peers).   But they described significant differences between the comparison groups on several individual MAYSI-2 items, not MAYSI-2 scales.

Finally, studies exploring the relation of substance use to antisocial characteristics have found ADU related to the *Psychopathic Personality Inventory's* second factor (Antisocial Personality) (r = .31: Toney-Smith et al., 2011) and to the Delinquent Behavior scale on the *Youth Self Report* (Grisso et al., 2001).

Relation to Past History Criteria

Suicide history.   In studies using two different samples, Archer et al. (2004, 2010) found that SI scores were significantly related to history of past suicide attempts and to independent interview data on youths' current suicidal thoughts. In the first study, using the Caution cut-point on SI produced very good sensitivity (.90) and specificity (.85) for predicting current suicidal thoughts, although sensitivity was not as high (.66) for identifying self-reported past suicide attempts. Wasserman et al. (2004a), however, found an AUC of .96 when using the SI Caution cut-point to identify youth reporting past suicide attempts on the *Diagnostic Interview Schedule for Children* (DISC).   Youth above the SI Caution cut-point were six times more likely to report lifetime suicide attempts than those below the cut-point.   However, results were not as good (AUC =. 55) when the SI Caution cut-point was used to identify youth high on DISC suicide ideation but *without* past attempts (Wasserman et al., 2004b).   Many youth reporting recent suicide thoughts on the DISC but *no* prior attempts did not score over the SI Caution cut-point.

A study of Austrian youth in juvenile custody (Plattner et al., 2007) used the Suicidality scale (self-reported suicidal thoughts and past attempts).in the *Mini-International Neuropsychiatric Interview for Children and Adolescents* (MINI-Kid: Sheehan et al., 2010). They entered all 7 MAYSI-2 scales and all 10 scales of the *Youth Self Report* into a signal

detention analytic model.   Among these 17 scales, the SI scale on MAYSI-2 was the best discriminator of youth high on MINI-Kid suicidality, followed by the YSR Withdrawal scale.

Trauma history. Archer et al. (2004) found no significant relation between TE (youths' reports of potential traumatizing past experiences) and records of youths' past abuse or neglect. But Archer et al. (2010) and Steele (2000) reported a significant relation between TE and records of girls' past sexual and physical abuse.   Caldwell et al. (2006) examined "affective states" (represented by DA and AI) and their relation to a measure of discomfort that focuses on youths' expressed symptoms in home and school environments. They found substantial correlations (most within the range of .25 and .60), but differences by race and gender. For example, for black males, DA scores were less related to symptoms experienced in home/school environment compared to white and Hispanic youth. Correlations generally were higher for girls than boys.

Mental health services history.   Steele (2000) reported no evidence that incarcerated youths' MAYSI-2 scores were related to past receipt of mental health services in the community. But Lopez-Williams et al. (2006) warned of a race/ethnicity effect that may mask the relation. In their study, the number of MAYSI-2 scales over the Warning cut-point was significantly higher for youth with greater past enrolment in mental health services when examined for whites, but black youth with less or more past mental health services did not differ on this MAYSI-2 index. Lopez-Williams noted that this race-related difference has been found in several other studies using other measures of psychopathology, suggesting race-related bias in mental health referral practices and/or race-related differential diagnostic error by clinicians making referrals without the benefit of measures of psychopathology. Finally, Archer et al. (2010) reported modest correlations between certain MAYSI-2 scales and having received certain psychiatric medications (e.g., DA and antidepressant medications).

Relation to Seriousness of Mental Health Problems

Two studies found no significant MAYSI-2 scale differences between youths with different degrees of impairment in functioning due to mental health problems (Bailey, 2008) and differences in severity of mental health problems based on background records (Vaughn et al., 2007, TE being an exception).   In contrast, in Cruise et al. (2011), youth who were designated "emotionally disturbed" in juvenile corrections assessments were more likely to be elevated on AI, DA and SI than other youths (although inclusion of the MAYSI-2 in the assessment that led to the "emotionally disturbed" designation might have influenced these results).   In another study, incarcerated youth divided into two MMPI-A clusters indicating greater and lesser general mental health problems were differentiated by higher DA and SI scores for the cluster group indicating greater disturbance (Espelage et al., 2003).   Finally, Newbauer and Stone (2010) found a significant correlation (r = -.50) between a "MAYSI-2 summary score" (the sum of scores on the MAYSI-2 scales) and a measure of general mental health (the *Sulliman Scale of Social Interest*, an Adlerian scale with a history of use in delinquent samples). The greatest relationships were with MAYSI-2 AI, DA and SI.

Of special interest is Cruise et al.'s (2008) use of mental health records to identify youth with and without "serious mental illness," then using these groups to identify a set of "MAYSI-2 critical items" that best differentiated between the groups.   Determining the clinical utility of this "MAYSI-2 Critical Items Scale" will require further work.

Relation of External Criteria to MAYSI-2 Classification Schemes

Some studies have used analytic procedures to create relatively homogeneous groups of young people based on patterns of MAYSI-2 scale scores, then examined characteristics of those groups.   Two of these studies (Cruise et al., 2007; Stewart & Trupin, 2003) independently used

K-means cluster analyses of MAYSI-2 scale scores in juvenile corrections samples, and both studies identified a "co-occurring" cluster, a cluster with high scores on scales other than ADU, and a group with generally lower scores.   Stewart and Trupin's youth with higher MAYSI-2 scores (the first two clusters) were more likely to have received mental health services prior to their incarceration.   Cruise et al. found significantly more sexual and physical abuse in past histories for the two more disturbed clusters.

Relation of Scales to Psychiatric Diagnoses

The MAYSI-2 was not designed to provide psychiatric diagnoses, and one would not expect it to be able to do so.   Its scales measure various symptom states, and many symptoms during adolescence are heterotypic—found in varying degrees in many diagnostic conditions—because psychopathology during adolescent development is characterized by less differentiation than in adulthood.   Nevertheless, MAYSI-2 scales identifying certain symptoms should at least be related (even if not substantially) to psychiatric diagnoses that include the symptom.

Wasserman et al. (2004a, 2004b) and Hayes et al. (2005) examined the relation between MAYSI-2 scales and diagnoses obtained with the *Voice-DISC* (a computer-assisted version of the *Diagnostic Interview Schedule for Children*).   In general, they found that when the MAYSI-2 scales are used together (e.g., over Caution cut-point on one or more scales), they can perform the task of identifying youth who are likely to receive *some* psychiatric diagnosis.   For example, Wasserman et al. (2004a) found that about 89% of youth who had a DISC diagnosis in a single diagnostic cluster (e.g., any affective disorder) scored over the Caution cut-point on at least one of the MAYSI-2 scales.   This increased to 97% of youth who qualified for DISC disorders in more than one diagnostic cluster.

But several studies have concluded that individual MAYSI-2 scales typically do not allow one to infer with accuracy that youth will meet criteria for a *specific* psychiatric diagnosis. For example, *ADU scores* above the Caution cut-point achieved good AUC coefficients of .77 (Wasserman et al., 2004a) and .80 (Hayes et al., 2005) for identifying a DISC diagnosis of any substance disorder, as well as a positive predictive value of .71 (Wasserman et al., 2004a). But Wasserman et al. (2004b) cautioned that 35% of youth meeting DISC criteria for substance use disorder scored below the ADU Caution cut-point. Both of those studies reported substantial relations between AI scores above the Caution cut-point and both disruptive and affective disorders defined by the DISC. But using AI scores to "predict" disruptive disorders was not supported; among youth with disruptive disorders, 43% scored below the AI Caution cut-point, and 57% of youth without DISC disruptive disorders scored above the AI Caution cut-point. This is consistent with the heterotypic nature of anger and irritability in adolescence; these emotional states are observed among youth with various disorders and with no disorders. Finally, the Wasserman et al. and Hayes et al. studies found AUC's of .70 and .80 for DA Caution cut-scores to identify youth with "any anxiety disorder" or "any affective disorder." Yet, while most youth with a DISC affective disorder scored above the DA Caution cut-point, up to 80% of youth scoring above the DA Caution cut-point did not meet criteria for an affective disorder.

The results would suggest that simplistic notions about how MAYSI-2 symptoms might identify specific diagnostic conditions should be avoided in clinical use. AI should not be used to predict disruptive behavior disorders, and DA does not predict DSM affective disorders. But either or both of them, when elevated, should call for a further assessment to determine the presence of possible disorders in which anger or depressed affect are commonly seen (for example, ADHD, depression and other mood disorders).

Relation of Scales to Justice System Processing

Several studies compared MAYSI-2 scale scores of groups that were formed by juvenile justice system decisions made during the legal processing of youths' cases.   Regarding youth who were diverted from adjudication or were processed for adjudication, Daurio et al. (2010) found that the latter were higher only on AI.   Stinson (2006) reported differences between diverted and processed youth on specific MAYSI-2 items.   Stewart and Trupin (2003) found that youth in MAYSI-2 clusters identified by greater mental health problems (with or without substance use) were more likely to receive more serious sentences.   Kalmbach (2006) reported that youth receiving greater sentences for the same offenses were significantly higher on ADU and TE.   Similarly, Marczyk et al. (2005) found higher ADU and TE scores among youth who were not decertified for trial in criminal court (that is, were retained in adult court for trial) compared to those who were.   Caution should be exercised in interpreting the results of these studies.   Their methods do not necessarily show that symptoms themselves directly influenced juvenile justice decisions.   Symptoms might co-vary with other decision-related characteristics not included as variables in the studies.

Discussion

The body of research on the MAYSI-2 is substantial despite its brief history.   This may be due, in part, to its use for screening all youth in a large number of juvenile justice settings nationwide.   This offers considerable potential for researchers to develop data sets sufficiently large for analysis of multiple demographic and setting variables when studying behavioral and mental health symptoms of young people entering the juvenile justice system.

With various limits and some exceptions, research using the MAYSI-2 scales has found adequate psychometric properties, internal consistency, and relations to other validated tools

measuring similar constructs.   Many studies demonstrate the relation of its scales to behavioral and mental health criteria theoretically associated with the constructs that the scales claim to measure.   Often the results have replicated patterns of psychopathology affirmed by studies using other methods (e.g., race/ethnicity-related patterns of substance use and patterns of mental health service referral).   These general observations across studies apply especially to the ADU, AI, and SI scales, and surprisingly to the TE scale.   They are least satisfactory for the TD scale for boys; research to strengthen that scale (and find a comparable construct for girls) could be valuable.

Cautions for Clinical Use

The review suggests several cautions for clinical application of the MAYSI-2.   The results affirm a limit associated with all screening tools—they should be used to identify youth for further assessment and mental health consultation, not for providing diagnoses or long-term treatment plans.   In addition, current studies only support the MAYSI-2's value when it is employed in the contexts for which it was intended: specifically, in juvenile justice programs, with youth soon after admission. Little is known about the meaning of MAYSI-2 information when used in non-juvenile-justice settings or at points in juvenile justice programs other than entry.

The relation of MAYSI-2 scores to race/ethnicity depends on contexts and specific scales.   MAYSI-2 studies tend to use non-selected samples of youth at intake to juvenile justice programs, and black and Hispanic young people constitute the majority in those settings in the U.S.   Thus MAYSI-2 norms and tests of hypotheses in perhaps the majority of MAYSI-2 studies are based primarily on young people of color.   Moreover, when data are analyzed by race/ethnicity, there are few consistent differences normatively other than ADU (substance use). But when studies examine MAYSI-2 scales' relations to other phenomena (e.g., referral to mental health services), the results sometimes suggest race-related differences in how the

MAYSI-2 is, or should be, interpreted. Further research (described later) could offer better guidance for adjusting clinical uses of MAYSI-2 data with various race/ethnic populations.

Suggestions for MAYSI-2 Research Methods

The review revealed that studies using the MAYSI-2 have employed a wide range of approaches for representing MAYSI-2 scores. Among them are the use of continuous scale scores, dichotomous scores based on the Caution or Warning cut points, trichotomous scores (below Caution, between Caution and Warning, above Warning), number of scales with scores over a cut point, the sum of scores on all scales, and item-by-item analysis. Different criteria are probably appropriate for different research purposes, but their diversity has some disadvantages. It makes comparisons between studies more difficult. In addition, researchers should be aware that results using continuous-scale analyses alone will be difficult to apply in clinical practice. Dichotomous or trichotomous formats (employing the standard cut points) are the only form of MAYSI-2 scores employed by juvenile justice personnel in everyday practice. If researchers' results are to be useful for clinical practice, they must be translated to allow for binary decision rules that juvenile justice personnel can apply.

MAYSI-2 data often are collected in the course of routine use of the MAYSI-2 with samples of young people consecutively admitted to juvenile justice programs. When differences in findings arise between MAYSI-2 studies, one must consider the possibility that these might reflect differences between the settings that are outside the study's formal variables and that are confounding direct interpretation of the differences in results.

These setting differences that can confound cross-study comparisons are of three kinds. Researchers who can report any of them in their Method sections will improve inferential comparisons to other sites. First, *administration variables* may differ considerably across

juvenile justice settings: for example, hours or days between admission and administration of MAYSI-2, adherence to standardized administration, and information provided to youth about uses of the results. Second, *population selection variables* include factors in a setting that operate to determine which young people enter a particular juvenile justice program. Such variables can make one community's pre-trial detention population somewhat different from another. Among these selection variables are laws and policies regarding detention criteria, police practices and decisions, and front-door diversion options. Especially influenced by these selection variables are cross-site differences in the race, gender, and offense-type of juvenile justice populations. Third, *hazard conditions* vary across communities, potentially influencing the level of psychopathology represented in their juvenile justice facilities. Among these are a variety of conditions known to influence mental health symptoms: for example, relative availability of mental health services, community environments that vary in exposing their youth to potentially traumatizing conditions, and different patterns and volume of drug traffic.

Perhaps because of such setting differences, studies often report different proportions of youth over Caution cut points than the proportions described in the MAYSI-2 original norms or National Norms. Discussion-section comments in research reports sometimes raise the possibility of adjusting the cut points to reflect "local" norms. Adjustments also are sometimes contemplated when certain demographic groups have been found consistently to be higher or lower than others: e.g., lowering the cut point on ADU for black youth, or raising the cut score on many MAYIS-2 scales for girls. This is ill-advised for both clinical and research purposes. In clinical use, it masks what may be true differences that call for attention to a youth's mental health needs. For research, the practice of adjusting cut points on MAYSI-2 scales would create

a body of research studies using different criteria and therefore creating difficulties associated with cross-study comparison.

Questions That Need MAYSI-2 Research Attention

The studies reported in this review offer many hypotheses for further investigation. Especially important are further efforts to improve the TD scale and to seek one for girls. Efforts to examine the relation between TE, other MAYSI-2 symptoms, and trauma-related conditions could improve the MAYSI-2's use in screening for PTSD. Beyond these suggestions, we offer several directions for future MAYSI-2 research in areas not yet explored or examined in preliminary ways by existing studies.

First, few of the existing studies have examined the utility of "decision rules" that juvenile justice programs apply in determining when a youth "screens in" for further attention. The most common MAYSI-2 decision rule was first developed by Pennsylvania's juvenile detention system: "Over Caution cut score on SI, or over Warning cut score on any other two scales." (For others, see National Center for Mental Health and Juvenile Justice, 2006). How do such decision rules perform in identifying youth who actually meet criteria for some mental disorder? Similarly, can statistically significant findings regarding correlations between MAYSI-2 scales and important behaviors (e.g., in-custody infractions) be translated into decision rules that juvenile justice staff can use? Can the weighted use of multiple MAYSI-2 scales better identify youth who obtain specific psychiatric diagnoses?

Second, several published studies point to the need to examine MAYSI-2 administration variables that might have a systematic influence on results. Some of these pertain to method: for example, time between admission and administration, and individual or group administration. Others refer to factors that might influence response style: for example, frequency and recency of

exposure to the MAYSI-2, and youths' perceptions of the manner in which staff will use the results.   Are there ways that response styles (e.g., under-reporting of problems) could be detected?

Third, no published studies have examined the fidelity and benefits of mental health screening with MAYSI-2.   When it is implemented in detention centers, does its use increase staff referral to mental health consultations? What is its impact on suicide prevention in custody?

Finally, some published studies suggest the need to examine the degree to which certain MAYSI-2 items or scales might have different latent meanings for boys and girls, or for various race/ethnic groups of youth.   DIF analysis (item response analysis) can serve this purpose. Typically, such analyses require large sample sizes.   But many MAYSI-2 studies can meet this requirement, because of the opportunity to use data collected across time in routine practice in one or more juvenile justice facilities.

In closing, it is likely that MAYSI-2 publications describing multi-national and multi-cultural results will appear more frequently in the future, as the result of the recent formation of a consortium of juvenile justice and mental health researchers in Europe. The consortium's researchers study behavioral and mental health symptoms among youth in juvenile justice settings in 8 countries.   As a group, they seek uniformity of methods and data collection across studies, including use of the MAYSI-2, to improve cross-study comparisons with the diverse cultural groups of young people found in juvenile justice settings in Europe.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders*

(4th ed., text rev.). Washington, DC: Author.

Archer, R., Sionds-Bisbee, E., Spiegel, D., Handel, R., & Elkins, D. (2010). Validity of the

Massachusetts Youth Screening Instrument-2 (MAYSI-2) in juvenile justice settings.

Journal of Personality Assessment, 92, 337-348. doi: 10.1080/00223891.2010.482009

Archer, R. P., Stredny, R. V., Mason, J. A., & Arnau, R. C. (2004). An examination and

replication of the- psychometric properties of the Massachusetts Youth Screening

Instrument - Second Edition (MAYSI-2) among adolescent in detention. *Assessment,*

11, 1-13. doi:10.1177/1073191104269863

Bailey, S.M. (2008). Predicting mental illness with the MAYSI-2. *Dissertation Abstracts*

International: Section B: The Sciences and Engineering, 68(9), 6287.

Bender, K., Perron, B., Howard, M., & Jenson, J. (2010). Strong-arm bullying behavior prior to

incarceration among a sample of young offenders. *Journal of the Society for Social*

Work and Research, *1*, 1-13.

Bisbee, E. (2010). Massachusetts Youth Screening Instrument long-term outcomes and scale

stability. *Dissertation Abstracts International: Section B: The Sciences and Engineering,*

71(1), 650.

Butler, M., Loney, B., & Kistner, J. (2007). The Massachusetts Youth Screening Instrument as a

predictor of institutional maladjustment in severe male juvenile offenders. *Criminal*

Justice and Behavior, *34*, 476-492. doi: 10.1177/0093854806291711

Chapman, J., & Ford, J. (2008). Relationships between suicide risk, traumatic experiences, and

    substance use among juvenile detainees. *Archives of Suicide Research*, *12*(1), 50-61.

    doi: 10.1080/13811110701800830

Caldwell, R., Sturges, S., & Silver, N. (2006). Home versus school environments and their

    influences on the affective and behavioral states of African-American, Hispanic, and

    Caucasian juvenile offenders. *Journal of Child and Family Studies, 16*, 125-138.

    doi: 10.1007/s10826-006-9073-6

Cauffman, E. (2004). A statewide screening of mental health symptoms among juvenile

    offenders in detention. *Journal of the American Academy of Child & Adolescent*

    Psychiatry, *43*(4), 430-439. doi:10.1097/00004583-200404000-00009

Cauffman, E., Lexcen, F., Goldweber, A, Shulman, E., & Grisso, T. (2007). Gender differences

    in mental health symptoms among delinquent and community youth. *Youth Violence and*

    Juvenile Justice, 5, 287-307. doi: 10.1177/1541204007301292

Cauffman, E., & MacIntosh, R. (2006). A Rasch differential item functioning analysis of the

    Massachusetts Youth Screening Instrument: Identifying race and gender differential Item

    functioning among juvenile offenders. *Educational and Psychological Measurement*,

    66, 502-521. doi:10.1177/0013164405282460

Cruise, K., Dandreaux, D., Marsee, M., DePrato, D. (2008). Identification of critical items on the

    Massachusetts Youth Screening Instrument-2 (MAYSI-2) in incarcerated youth. *The*

    International Journal of Forensic Mental Health, 7, 121-132.

    doi: 10.1080/14999013.2008.9914409

Cruise, K., Evans, L., & Pickens, I. (2011). Integrating mental health and special education needs into comprehensive service planning for juvenile offenders in long-term custody settings. *Learning and Individual Differences*, *21*, 30–40. doi: 10.1016/j.lindif.2010.11.004

Cruise, K., Marsee, M., Dandreaux, D., & DePrato, D. (2007). Mental health screening of female juvenile offenders: Replication of a subtyping strategy. *Journal of Child and Family Studies*, *16*, 615-625.   doi: 10.1007/s10826-006-9111-4

Dalton, R. , Evans, L., Cruise, K., Feinstein, R., & Kendrick, R. (2009) . Race differences in mental health access in a secure male juvenile justice facility. *Journal of Offender Rehabilitation, 48*, 194-209. doi: 10.1080/10509670902766570

Daurio, R. (2010). Factors associated with court decisions to provide juvenile offenders with mental health placements. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 70*(8), 5156.

DeLisi, M., Drury, A., Kosloski, A., Caudill, J., Conis, P., Anderson, C., Vaughn, M., & Beaver, K. (2010b). The cycle of violence behind bars: Traumatization and institutional misconduct among juvenile delinquents in confinement. *Youth Violence and Juvenile Justice, 8*(2), 107-121. doi: 10.1177/1541204009349399

DeLisi, M., Caudill, J., Trulson, C., Marquart, J., Vaugh, M., & Beaver, K. (2010a). Angry inmates are violent inmates: A Poisson regression approach to youthful offenders. *Journal of Forensic Psychology Practice*, *10*, 419-439.
doi: 10.1080/15228932.2010.489861

Espelage, D., Cauffman, E., Broidy, L., Piquero, A. R., Mazerolle, P., & Steiner, H. (2003). A cluster-analytic investigation of MMPI profiles of serious male and female juvenile offenders. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*,

770-777. doi: 10.1097/01.CHI.0000046877.27264.F6

Ford, J., Chapman, J., Pearson, G., Borum, R., & Wolpaw, J. (2007). Psychometric status and

clinical utility of the MAYSI-2 with girls and boys in juvenile detention. *Journal of

Psychopathology and Behavioral Assessment*, *30*, 87-99. doi:10.1007/s10862-007-9058-9

Garland, E, & Howard, M (2010a). Inhalation of computer duster spray among adolescents: An

emerging public health threat? *The American Journal of Drug and Alcohol Abuse, 36*,

320-324. doi: 10.3109/00952990.2010.504874

Garland, E, & Howard, M. (2010b). Phenomenology of adolescent inhalant intoxication.

Experimental and Clinical Psychopharmacology, 18(6), 498-509. doi: 10.1037/a0021737

Gretton, H., & Clift, R. (2011). The mental health needs of incarcerated youth in British

Columbia, Canada. *International Journal of Law and Psychiatry, 34*, 109-115.

doi: http://dx.doi.org/10.1016/j.ijlp.2011.02.004

Grisso, T. & Barnum, R. (2000). *Massachusetts Youth Screening Instrument-Version 2: User's

Manual and Technical Report*. Worcester, MA: University of Massachusetts Medical

School.

Grisso, T. & Barnum, R. (2006). *Massachusetts Youth Screening Instrument-Version 2: User's

Manual and Technical Report*. Sarasota, FL: Professional Resource Press.

Grisso, T., Barnum, R., Fletcher, K. E., Cauffman, E., & Peuschold, D. (2001). Massachusetts

Youth Screening Instrument for Mental Health Needs of Juvenile Justice Youths. *Journal

of the American Academy of Child & Adolescent Psychiatry*, *40*(5), 541-548.

doi:10.1097/00004583-200105000-00013

Grisso, T., & Quinlan, J. (2005). Massachusetts Youth Screening Instrument-Revised. In T. Grisso, G. Vincent, & D. Seagrave (eds.), *Mental health screening and assessment in juvenile justice* (pp. 99-111). New York: Guilford Press

Grisso, T., Vincent, G., & Seagrave, D. (2005). *Mental health screening and assessment in juvenile justice*. New York: Guilford Press

Harshbarger, J. (2007). The impact of mental health dimensions on the prediction of juvenile reentry recidivism. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 67*(10), 6057.

Hayes, M., McReynolds, L., & Wasserman, G. (2005). Paper and voice MAYSI-2: Format comparability and concordance with the Voice DISC-IV. *Assessment*, *12*, 395-403. doi:10.1177/1073191105280359

Hickle, K., & Roe-Sepowitz, D. (2010). Female juvenile arsonists: An exploratory look at c haracteristics and solo and group arson offences. *Legal and Criminological Psychology, 15*, 385-399. doi: 10.1348/135532509X473913

Johnson, D. (2006). The latent structure of adolescent problematic substance use: Taxometric analysis of the MAYSI-2 alcohol and drug use scale. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 67*(1), 547.

Kalmbach, K. (2006). Mental health and psychosocial characteristics of juvenile offenders: Relationship to dispositional outcomes. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 66*(9), 5092.

Kerig, P., Ward, R., Vandersee, K., & Moeddel, M. (2009). Posttraumatic stress as a mediator of the relationship between trauma and mental heatlh problems among juvenile delinquents. *Journal of Youth and Adolescence, 38,* 1214-1225. doi: 10.1007/s10964-008-9332-5

Logan, L. (2007). A cross-time analysis of the mental health needs of juvenile offenders

      Dissertation Abstracts International: Section B: The Sciences and Engineering, 68(5), 3402.

Lopez-Williams, A., Vander Stoep, A., Kuo, A., & Stewart, D. (2006). Predictors of mental

      health service enrollment among juvenile offenders. *Youth Violence and Juvenile Justice,*

      4(3), 266-280. doi: 10.1177/1541204006290159

Marczyk, G. , Heilbrun, K., Lander, T., & DeMatteo, D. (2003). Predicting juvenile recidivism

      with the PCL:YV, MAYSI, and YLS/CMI. *International Journal of Forensic Mental*

      Health, 2(1), 7-18.

Marczyk, G., Heilbrun, K., Lander, T., & DeMatteo, D. (2005). Juvenile decertification:

      Developing a model for classification and prediction. *Criminal Justice and Behavior, 32*,

      278-301.   doi: 10.1177/0093854804274371

Maney, S. (2011). Violent and nonviolent juvenile offenders: Comparisons concerning self-

      report of mental health symptoms and response styles (Unpublished doctoral

      dissertation). Massachusetts School of Professional Psychology, Boston, MA.

McConihay, J. (2008). Accuracy of detained juveniles self report of substance use on the

      Massachusetts Youth Screening Instrument-Version 2. *Dissertation Abstracts*

      International: Section B: The Sciences and Engineering, 68(12), 8444.

McCoy, H. (2010). Using cognitive interviewing to explore causes for racial differences on the

      MAYSI-2. *Crime & Delinquency.* Advance online publication. doi:

      10.1177/0011128710388922

McCoy, H**.** (2011). A path analysis of factors influencing racial differences on the MAYSI-2.

      Journal of Offender Rehabilitation, 50, 119-141. doi: 10.1080/10509674.2011.560549

National Center for Mental Health and Juvenile Justice (2006). *Mental health screening in juvenile justice: The next frontier.* Delmar, NY: National Center for Mental Health and Juvenile Justice.

Murrie, D., Henderson, C., Vincent, G., Rockett, J., & Mundt, C. (2009). Psychiatric symptoms among juveniles incarcerated in adult prison. *Psychiatric Services, 60*, 1092-1097. doi: 10.1176/appi.ps.60.8.1092

Newbauer, J., & Stone, M. (2010). Social interest and self-reported distress in a delinquent sample: Application of the SSSI and the MAYSI-2. *The Journal of Individual Psychology, 66*(2), 201-215.

Nordness, P., Grummert, M., Banks, D., Schindler, M., Moss, M., Gallagher, K., & Epstein, M. (2002). Screening the mental health needs of youths in juvenile detention. *Juvenile and Family Court Journal*, *53(2)*, 43-50. doi: 10.1111/j.1755-6988.2002.tb00061.x

Osterlind, S., Koller, J., & Morris, E. (2007). Incidence and practical issues of mental health for s school-aged youth in juvenile justice detention. *Journal of Correctional Health Care*, *13*, 268-277. doi:10.1177/1078345807306802

Padilla, R. (2006). Juvenile offenders' affective responses to a traumatic experience: A psychometric study using the Massachusetts Youth Screening Instrument-Version 2. *Dissertation Abstracts International: Section B: The Sciences and Engineering 67*(6)**.**

Plattner, B., The, S., Kraemer, H., Williams, R., Bauer, S., Kindler, J., Feucht, M., Friedrich, M., & Steiner, H. (2007). Suicidality, psychopathology, and gender in incarcerated adolescents in Austria. *Journal of Clinical Psychiatry, 68*(10), 1593-1600. Retrieved from http://www.psychiatrist.com/privatepdf/2007/v68n10/v68n1019.pdf

Roe-Sepowitz, D. (2007). Adolescent female murderers: Characteristics and treatment implications. *American Journal of Orthopsychiatry, 77*(3), 489-496. doi: 10.1037/0002-9432.77.3.489

Sheeber, Lisa B., Allen, N., Leve, C., Davis, B., Shortt, J., & Katz, L. (2009). Dynamics of affective experience and behavior in depressed adolescents. *Journal of Child Psychology and Psychiatry, 50*(11), 1419-1427. doi:10.1111/j.1469-7610.2009.02148.x

Sheehan, D., Sheehan, K., Shytle, R., Janavs, J., Bannon, Y., Rogers, J., Milo, K., Stock, S., & Wilkinson, B. (2010). Reliability and validity of the Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID). *Journal of Clinical Psychiatry, 71*(3), 313-26. doi: 10.1176/appi.ps.54.3.377

Stathis, S., Letters, P., Doolan, I., Fleming, R., Heath, K., Arnett, A., and Cory, S. (2008). Use of the Massachusetts Youth Screening Instrument to assess mental health problems in young people within an Australian youth detention centre. *Journal of Pediatrics and Child Health, 44,* 438–443. doi:10.1111/j.1440-1754.2008.01324.x

Steele, K. (2000). Dual-symptomed juvenile offenders: A study of mental health and substance abuse symptoms within Washington state juvenile rehabilitation administration. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 61*(6), 2199.

Stewart, D., & Trupin, E. (2003). Clinical utility and policy implication of a statewide mental health screening process for juvenile offenders. *Psychiatric Services, 54*(3), 377-382. doi: 10.1176/appi.ps.54.3.377

Stinson, P. (2006). Determining the prevalence of mental health needs in the juvenile

    delinquency system at intake: A MAYSI-2 comparison of non-detained and detained

    youth (Doctoral dissertation). Retrieved http://search.proquest.com/docview/305353326

Tille, J., & Rose, J. (2007). Emotional and behavioral problems of 13-to-18-year-old incarcerated

    female first-time offenders and recidivists. *Youth Violence and Juvenile Justice, 5,*

    426-435. doi: 10.1177/1541204007300355

Toney-Smith, S., Edens, J. F., & Vaghn, M. G. (2011). Assessing the external correlates of the

    alternative factor models of the psychopathic personality inventory – short form across

    three samples. *Journal of Personality Assessment, 93*, 244-256.

    doi: 10.1080/00223891.2011.558876

Vance, D. (2006). The relationship between scores on the Massachusetts Youth Screening

    Instrument and serious incidents committed by juveniles detained in the Dallas County

    Juvenile Detention Center (Texas). *Dissertation Abstracts International Section A:*

    Humanities and Social Sciences, 66(10), 3578.

Vaughn, M., Freedenthal, S., Jenson, J., & Howard, M., (2007). Psychiatric symptoms

    and substance use among juvenile offenders: a latent profile investigation. *Criminal*

    Justice and Behavior, 34(10), 1296-1312. doi: 10.1177/0093854807304624

Vaughn, M., Howard, M., & Chang-Harper, L. (2006). Do prior trauma and victimization predict

    weapon carrying among delinquent youth? *Youth Violence and Juvenile Justice, 4*(4), 314-

    327. doi: 10.1177/1541204006292665

Vincent, G., Grisso, T., Terry, A., & Banks, S. (2008). Sex and race differences in mental

    health symptoms in juvenile justice: The MAYSI-2 national meta-analysis. *Journal of*

    the American Academy of Child and Adolescent Psychiatry, 47, 282-290.

doi: 10.1097/CHI.0b013e318160d516

Wasserman, G. , McReynolds, L., Ko, S. , Katz, L. , Cauffman, E., Haxton, W., Lucas, C.

(2004a). Screening for emergent risk and service needs among incarcerated youth:

comparing the MAYSI-2 and voice DISC-IV. . *Journal of American Academy of Child*

and Adolescent Psychiatry, 43, 629-639. doi: 10.1097/01.chi.0000116742.71662.b1

Wasserman, G., Vilhauer, J., McReynolds, L., Shoai, R., & John, R. (2004b). Mental health

screening in the juvenile justice system: A comparison between the Voice-DISC-IV and

the MAYSI-2. *Journal for Juvenile Justice Services 19,* 7-17.

Appendix

Description of MAYSI-2

The *Massachusetts Youth Screening Instrument –Version* 2 (MAYSI-2) was designed

specifically as a screening (triage) tool to be administered to all youth (ages 12 through 17),

often by non-clinical personnel, at the time of intake (within 1-3 hours after admission) in

juvenile probation offices, juvenile pretrial detention centers, and juvenile corrections facilities

(Grisso et al., 2001). Its purpose is to identify symptoms (thoughts, feelings and behaviors) that

are found in many psychiatric diagnostic conditions of youth, but it was not intended to provide

diagnoses. Ideally it offers information to intake staff that might require an immediate mental

health response (e.g., suicide precautions, need for further evaluation, referral for clinical

consultation), but it was not developed to provide treatment recommendations.

History

The original studies that created the MAYSI prototype were performed between 1995 and

1998, at the University of Massachusetts Medical School, with a grant from the William T. Grant

Foundation and in collaboration with the Massachusetts Department of Youth Services. Those

studies involved item selection, examination of psychometric properties and initial validity, and

creating cut points. The original MAYSI was released in 1998 for use in California,

Massachusetts and Pennsylvania in order to accumulate further data to refine the tool

psychometrically. Analyses of those data (Grisso et al., 2001; Grisso & Barnum, 2000) resulted

in the current 7 scales for the second version (MAYSI-2), which was published and released for

use in 2000. The John D. and Catherine T. MacArthur Foundation provided support (and

continues to do so) for a MAYSI-2 technical assistance and research center at the University of

Massachusetts Medical School, called the *National Youth Screening and Assessment Project*

(NYSAP).   The third project to refine the MAYSI-2, called the MAYSI-2 National Norms Project, involved collecting over 70,000 cases nationwide (Grisso & Barnum, 2006: Manual National Norms Supplement; see also Vincent et al., 2008). This produced new national norms but did not suggest the need for re-scaling to create a third version.

Items and Scales

The 52 items in the MAYSI-2 questionnaire are questions about behaviors, thoughts or feelings that young people answer "yes" or "no" as being true for them in the "past few months." Administration is performed individually using paper-and-pencil or computer-assisted administration (using software called MAYSIWARE).   The items are brief and worded simply: for example, "Have you had a lot of trouble falling asleep or staying asleep?" "Have you felt that life was not worth living?" "Have you felt angry a lot?" With the paper-and-pencil administration, young people read the questions to themselves (requiring about a fifth-grade reading level) or, if they cannot read well, the items are read by the staff member while the youth circles the answers for each item.   MAYSIWARE reads each item aloud to the youth through headphones as it appears on the screen. Administration and scoring require about 15 minutes.

The prototype MAYSI (pre-2000) had 9 scales formed on conceptual grounds.   Factor analyses, however, consistently identified 7 scales, and the 7-scale version became MAYSI-2 (2000 forward).   Factor analyses revealed only 6 stable scales for girls, so that the Thought Disturbance scale is scored only for boys.   Each scale (see Table 1 in the text of this article) is composed of 5 to 9 items.   Some items contribute to more than one scale. A few items in the MAYSI-2 contribute to no scale.   This is because ihey did not contribute significantly to the factors that arose in the above analyses, but they were retained for potential use in future research.

Scoring and Decision Aids

Every item answered "yes" adds a point to the scale to which it belongs. The MAYSI-2 does not provide a total score. Continuous scale scores and means are sometimes used in MAYSI-2 research, but the MAYSI-2 employs standardized cut points that were developed for clinical use to express low/high scores dichotomously. The MAYSI-2 norms provide two types of cut points: *Caution* and *Warning* (the latter being higher). As described in detail in the MAYSI-2 manual, the standardized *Caution* cut points were selected based on concurrent administration of other assessment tools (the *Millon Adolescent Clinical Inventory* and the *Youth Self Report/Child Behavior Checklist*) with known validity for identifying youth with clinically-significant levels of the symptoms in question. ROC analyses were used to set the Caution cut points on the MAYSI-2 scales so that they operated as similarly as possible to the clinical cut points on the concurrent, validated assessment instruments. The *Warning* cut points were set to identify the group of youth who scored in the top 10% of youth on the MAYSI-2 scale in question in the original normative samples. Caution and Warning cut points are provided for 6 of the scales, the exception being *Traumatic Experiences* (TE). Unlike the other scales, TE scores indicate self-reported exposure to events, not symptoms.

Individual juvenile justice programs develop their own policies regarding the manner in which the cut points are used to identify cases requiring immediate attention. For example, one of the more frequently-used decision rules was developed in Pennsylvania: "Over Caution cut point on Suicide Ideation, or over Warning cut point on any two scales." This decision rule typically identifies about 20% of a juvenile detention center's admissions as needing follow-up to screening.

Conditions of Use

The MAYSI-2 is not in the public domain.   It is published by Professional Resource Press, and copyright is held by the University of Massachusetts Medical School. A MAYSI-2 research and technical assistance center is operated at UMMS, called the *National Youth Screening and Assessment Project* (NYSAP), offering a variety of services to MAYSI-2 users.   Purchase of the MAYSI-2 manual and MAYSIWARE (the MAYSI-2 software) are unrestricted, but permission to use the materials in juvenile justice programs requires purchase of a manual for each facility and submission of a registration form (enclosed in the manual) to NYSAP for each facility.   A return copy with NYSAP's signature authorizes the facility to administer the MAYSI-2 to youth with no per-case fees.   Neither the MAYSI-2 manual nor the MAYSI-2 questionnaire items may be posted on the internet.

Table 1: MAYSI-2 Scales and Caution/Warning Cut-Scores

Acronym	Scale Name	Number Of Items	Brief Description
ADU	Alcohol/Drug Use	8	Frequency and pervasiveness of use of substances
AI	Angry-Irritable	9	Feelings of preoccupying anger and vengefulness, irritability and "touchiness"
DA	Depressed-Anxious	9	Depressed and/or anxious feelings
SC	Somatic Complaints	6	Bodily aches and pains often related to depressed or anxious feelings
SI	Suicide Ideation	5	Thoughts and intentions about self-harm, feelings of hopelessness
TD	Thought Disturbance	5	Altered perceptions of reality, things not seeming "real"
TE	Traumatic Experiences	5	Self-reported exposure to events that have potential traumatizing effects

Caution and Warning Cut Pointss (provided for all scales except TE)

Caution Range: *Criterion:*   For each scale, all scores above a cut point established in the original development of the MAYSI-2 as the score best identifying youth scoring in the "clinical range" on other validated assessment instruments measuring a similar symptom construct

Interpretation: Youth has scored at a level that may have "clinical significance"

Warning Range: *Criterion*: For each scale, all scores above a cut point established as the point that identified the highest 10% of youths on the original MAYSI-2 normative samples

Interpretation: Youth has scored exceptionally high in comparison to other youths in juvenile justice settings.