Published in final edited form as:

Assessment. 2005 March ; 12(1): 28–39. doi:10.1177/1073191104270838.

Ability of Substance Abusers to Escape Detection on the Minnesota Multiphasic Personality Inventory–Adolescent (MMPI-A) in a Juvenile Correctional Facility

L. A. R. Stein and

Brown University; The Rhode Island Training School

John R. Graham

Kent State University

Abstract

The ability of respondents to underreport successfully on substance abuse and validity scales of the Minnesota Multiphasic Personality Inventory–Adolescent was evaluated. Incarcerated teens (67 substance abusing, 59 non-substance abusing) completed the MMPI-A twice: once under standard instructions (SI) and once under instructions to fake good (FG). Under SI, substance scales correctly classified about 60% to 85% of adolescents. Under FG, substance- and non-substance-abusing juveniles produced lower scores on substance scales. However, the Lie Scale (L) was able to detect more than 75% of deceptive profiles and about 77% of honest profiles. When scale L and the best substance scale were used in combination, only about 18% of faking substance abusers were not identified as either substance abusers or as underreporting. For feigning substance abusers, only about 10% of substance abusers were detected, with about 72% being categorized as faking and needing further assessment.

Keywords

incarcerated; faking; MMPI-A; substance use

Given the social and economic costs of adolescent substance abuse, professionals have attempted to develop valid and efficient screening and assessment instruments to identify adolescents with such problems (Personal Experiences Inventory [PEI], Winters, 1999; Substance Abuse Subtle Screening Inventory for Adolescents [SASSI-A], Miller, 1990). It is especially important to screen and assess for such disorders in juvenile correctional settings, where often this is the adolescent's first opportunity for intervention. There is a high rate of substance use and abuse in justice-involved teens (Teplin, Abram, McClelland, Dulcan, & Mericle, 2002), and it is important to be able to identify those adolescents who need more services and those who may not require as many services from a system with few resources.

There has been increasing interest in the application of the MMPI-A (Butcher et al., 1992) to juvenile offender populations (e.g., see Cashel, Rogers, Sewell, & Holliman, 1998; Pena, Megargee, & Brody, 1996; Stein & Graham, 1999, 2001). Screening and assessing for substance problems can be important for understanding factors that may have been involved in misconduct as well as treatment placement. However, screening and assessment can be impeded by motivation to misreport symptoms of substance use. It is important that

screening and assessment measures be able to effectively detect substance use difficulties and misreporting. Fortunately, the MMPI-A has scales to detect substance use problems as well as misreported symptoms and behavior.

Detecting Substance Abuse Using the MMPI-A

The MMPI-A has three scales to detect substance use problems: the MacAndrew Alcoholism Scale-Revised (MAC-R; Butcher et al., 1992), the Alcohol/Drug Problem Acknowledgment (ACK) scale, and the Alcohol/Drug Problem Proneness (PRO) scale (Weed, Butcher, & Williams, 1994). The MAC-R was originally devised by MacAndrew in 1965 and was called the MAC. Effectiveness of the MAC and MAC-R have been reviewed elsewhere (see Stein & Graham, 2001), and generally, results have been mixed at best. Conduct-disordered adolescents appear to achieve high MAC scores regardless of substance abuse status (Wolfson & Erbaugh, 1984). The use of MAC in nonclinical, psychiatric, and substance-abusing adolescents also was studied (Gantner, Graham, & Archer, 1992), and an overall hit rate of approximately 77% was obtained when classifying nonclinical and substance-abusing adolescents. However, rates were reduced to approximately 65% when substance abusers were compared with psychiatric adolescents, of whom almost 30% were conduct-disordered (Gantner et al., 1992). Stein and Graham (2001) studied the ability of the MAC-R to detect substance abuse in a juvenile correctional setting and found that this scale was unrelated to substance use status.

The efficacy of MMPI-A substance abuse scales in detecting substance abuse has been understudied. A rational-empirical approach was used to develop ACK and, as a result, this is a more face-valid measure of substance use than the empirically derived PRO (see Weed et al., 1994). In one study, PRO and ACK were found to discriminate between substance-abusing adolescents and both clinical as well as nonclinical adolescents (Weed et al., 1994). In comparison to ACK and MAC-R, PRO appeared to perform best in making these distinctions (Weed et al., 1994). In discriminating between substance-abusing and clinical adolescents, MAC-R performed more poorly in comparison to PRO and ACK (Weed et al., 1994). Finally, Weed et al. (1994) compared clinical adolescents with and without substance use problems and found that PRO discriminated best between groups, whereas MAC-R performed more poorly than both PRO and ACK. T score cutoffs of 60, 65, and 70 have been suggested for ACK and PRO (Archer, 1997; Butcher et al., 1992; Weed et al., 1994).

PRO contains items relating to antisocial behavior (negative peer influences, limited involvement with parents, misbehaviors, and low achievement). Given the concordance between such behavior and substance abuse, this strategy of capitalizing on delinquency to identify substance abusers may be ineffective in juvenile correctional settings (Rogers & Kelly, 1997). Similarly, a number of investigators (e.g., Basham, 1992; Gottesman & Prescott, 1989; Wolfson & Erbaugh, 1984) have indicated that the MAC scale does not directly assess substance abuse and is most sensitive to antisocial and delinquent behaviors. On the other hand, about half of the ACK items have content that relates directly to substance use, and not antisocial or delinquent behavior in general. Whereas PRO and MAC-R are significantly associated with both substance abuse and behavioral under-control, ACK appears to be a somewhat specific measure of an adolescent's willingness to acknowledge substance abuse (Gallucci, 1997). In a recent investigation (Stein & Graham, 2001), the effectiveness of MAC-R, PRO, and ACK in distinguishing between substance-abusing and non-substance-abusing incarcerated adolescents was determined. Optimal cutoffs on PRO and ACK were at T score ≥ 60 and ≥ 55 , respectively; MAC-R was unrelated to substance use (Stein & Graham, 2001). Generally, ACK performed best in classifying participants with sensitivity as high as approximately 83% (Stein & Graham, 2001).

Detecting Underreporting Using the MMPI-A

In a relatively recent publication (Stein & Graham, 1999), detection of underreported symptoms and behaviors using the MMPI (Hathaway & McKinley, 1940), MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989), and MMPI-A was reviewed. As summarized in Stein and Graham (1999), generally, both traditional indicators of faking good (such as L, K, K – F, and L + K) as well as less familiar indicators (such as the Positive Malingering scale [Mp; Cofer, Chance, & Judson, 1949], Wiggins's Social Desirability scale [Wsd; Wiggins, 1959], Edwards's Social Desirability scale [Esd; Edwards, 1957], and the Superlative scale [S; Butcher & Han, 1995]) have shown efficacy in the detection of faking good. In general, scale L appears to be very effective in detecting underreporting of symptoms in adults.

Detection of faking good by adolescents is relatively understudied. Herkov, Archer, and Gordon (1991) presented cutoff scores for adolescents on various indexes of faking. The optimal L scale cutoff differentiating participants from clinical settings with fake-good instructions from participants found in public school settings with standard instructions was T score ≥ 55 (raw score ≥ 6). This score correctly classified 71.4% and 90.4% of participants with fake-good and standard instructions. Baer, Ballenger, and Kroll (1998) compared the profiles of clinical adolescents instructed to fake good with the profiles of both clinical and nonclinical adolescents who received standard instructions. Their findings support the continued use of scales L and K in detecting underreporting. T score cutoffs ranged from 57 to 67 on L and K, and correct classification rates ranged from approximately 62% to 91%. Overall, results suggest that the L scale appears to be the validity scale that best discriminates between fake-good profiles and profiles obtained under standard instructions for adolescents. Stein and Graham (1999) studied the ability of several MMPI-A indicators (L, K, L + K, K – F, Wsd, Mp, S, Esd) to detect faking good in incarcerated juveniles and found scale L to be the best indicator, classifying approximately 74% to 83% of teens correctly.

Detecting Underreported Substance Use

The importance of being able to detect misreports, specifically underreporting, during substance abuse evaluations is widely recognized (Rogers & Kelly, 1997; Stein et al., 2002). Perhaps this is why relatively recently, adolescent measures have begun to include faking scales in their designs (e.g., PEI, SASSI-A). Rogers and Kelly (1997) reviewed validity studies of the PEI and concluded that the rationale for declaring adolescent protocols as invalid is not explicit, and adequate cutoff scores have not been developed. Similarly, these authors reviewed studies on the ability of the SASSI to detect underreporting in adults (Rogers & Kelly, 1997) and concluded that more research is needed in establishing this test's ability to effectively detect underreporting in substance abusers. The SASSI-A has been studied in juvenile delinquents, where it correctly identified about 76% of adolescents underreporting substance use but incorrectly classified about 68% of non-substance abusers as abusers (Rogers, Cashel, Johansen, Sewell, & Gonzalez, 1997). Furthermore, these authors found evidence that the SASSI-A was ethnically biased (Rogers et al., 1997).

Evidence indicates that the MMPI-2 and MMPI-A are not ethnically biased (McNulty, Graham, Ben-Porath, & Stein, 1997; Stein, McClinton, & Graham, 1995). Furthermore, the MMPI, MMPI-2, and MMPI-A are among the most widely used questionnaires and few instruments have validity scales as well researched and with as much support for their use. Otto, Lang, Megargee, and Rosenblatt (1988) conducted one of the few studies to date that has rigorously examined detection of underreported substance use. These authors (Otto et al., 1988) studied the ability of alcohol-dependent inpatients and general medical patients to

underreport successfully on substance abuse and validity scales of the MMPI (Hathaway & McKinley, 1940). Under standard instructions (SI), correct classification was as high as 80% for alcohol problems, but under instructions to fake good (FG), profiles between the two groups were generally indistinguishable (Otto et al., 1988). In addition, faking scales were able to detect as much as 80% of invalid profiles under the FG condition (Otto et al., 1988). When the best validity and substance use indicators were used in combination, only 7.5% of faking alcoholics were not identified as either having an alcohol problem or as underreporting.

Because incarcerated teens may be motivated to avoid detection of substance abusive behaviors, the purpose of the current study is to determine the ability of the MMPI-A to distinguish between substance-abusing and non-substance-abusing incarcerated adolescents taking the test under SI and FG conditions ($N = 126$). It was hypothesized that despite the subtlety of the PRO scale, a significant proportion of substance abusers would be able to avoid detection. Furthermore, it was predicted that although a number of substance abusers would be able to avoid detection by ACK and PRO, the majority of their profiles would be identified as invalid by the L scale. Because MAC-R has been shown to be unrelated to substance use behaviors in this setting (Stein & Graham, 2001), this scale was not included in the current study.

This study is an extension of our work in detecting faking good (Stein & Graham, 1999) and in detecting substance abuse (Stein & Graham, 2001) in incarcerated adolescents. We utilize data reported in these previous studies. The unique contribution of this study lies in its examination of how well the MMPI-A detects substance abusers who are also faking good. The previous studies did not address this important issue.

METHOD

Participants

The sample ($N = 126$, see below) was collected in the Northeast section of the United States. The average age for participants was 15.33 years ($SD = 0.96$); 71.4% were boys; and 61.9% were African American, 4.8% were Hispanic, 1.6% were Native American, 31.0% were White, and 0.8% were self-described as "Other." Adolescents had their dispositional hearings within the juvenile justice system and were deemed to be in need of mental health treatment and intervention. Given their legal infractions, it was also determined that they were in need of a highly structured environment provided by both security officers and mental health professionals.

Throughout a 10-month period, of the 236 adolescents who entered the facility, 44 did not complete both the SI and FG sessions (due to teen refusal, behavioral difficulty, escape, facility discharge). Of the 192 remaining, 27 participants were eliminated because they did not understand or did not follow the directions (as determined by the Compliance Form, described later). Of the 165 remaining, 11 were eliminated because they were not between the ages of 14 and 18 years old, inclusive (as recommended by the MMPI-A manual; Butcher et al., 1992). Of the 154 adolescents remaining, 23 were eliminated on the basis of the MMPI-A exclusion criteria outlined below, leaving 131 participants. Of the 131 participants, 5 did not have substance abuse intake ratings recorded in their files (see Measures section below), leaving a final sample of 126 boys and girls.

As incentive to participate, adolescents received a TV and VCR for their newly developed Student Union Building. Adolescents received small snacks during testing and, if they completed all phases of the study, they received a bag of snack foods immediately following completion of testing. In addition, they were told that a prize (a board game or a party)

would be awarded to the cottage of the adolescent who did the best job following the directions. This sample was part of a previously reported study (Stein & Graham, 1999).

Instruments

Biographical information form—This form was constructed during development of the MMPI-A and included demographic information such as ethnic origin and age.

MMPI-A—The scales of the MMPI-A have adequate reliability and validity, and detailed information on the development and psychometric properties of this instrument can be found in Butcher et al. (1992). As found in the MMPI-A manual (Butcher et al., 1992), test-retest (1 week) reliability coefficients for nonclinical adolescents are as follows: L, $r = 0.61$; ACK, $r = 0.56$; and PRO, $r = 0.70$. No test-retest data have been reported in clinical samples. For combined nonclinical, clinical, and substance abuse samples, coefficient α for ACK is 0.76 and for PRO coefficient α is 0.70 (Weed et al., 1994). As reported in the MMPI-A manual (Butcher et al., 1992), coefficient α is about 0.61 in nonclinical boys and girls and about 0.54 in clinical boys and girls for scale L.

Compliance Questionnaire—The MMPI-A was administered under fake-good and standard instructions using a counterbalanced design (see Procedures for details). The Compliance Questionnaire asked participants to indicate on which testing session (first or second) they were asked to respond honestly and on which session they were asked to underreport symptoms and behaviors. This questionnaire also asked participants to indicate what they had done when asked to respond to items under standard and fake-good instructions (answered honestly, underreported symptoms and behaviors a little, underreported symptoms and behaviors a lot, overreported symptoms and behaviors).

Alcohol and Drug Use Survey (ADUS)—Clinical records were reviewed to obtain ADUS scores. Each participant was interviewed at intake by a Certified Chemical Dependence Counselor (CCDC) at the facility using the ADUS (Sweeney, 1987). This is a 15-item instrument with between 6 to 8 fixed-response choices. Items cover amounts, frequencies, and types of substances used; age of first use; with whom and why an adolescent uses; effects of using substances; and whether the adolescent or others have worried about the adolescent's use of alcohol or drugs. Although they did not have access to MMPI-A results, CCDCs had access to collateral data to guide and inform the ADUS interview. Collateral data available included family reports of teen behaviors and court records of crimes and mental health reports. ADUS scores ranging from 0 to 21 suggest minimal involvement, 22 to 50 indicate an adolescent uses substances, 51 to 83 indicate abuse of substances, and 84 to 147 suggest a dependency diagnosis (Sweeney, 1987). Although there are no published data on this survey, the usefulness of these scores in placing and treating youths was verified by the Ohio Department of Youth Services using record review and other interview data (Hawkins, 1987).¹

Procedures

Informed consent—For each youth, written, informed consent was obtained from the guardian and from the adolescent prior to participation.

Test administration—The taped version of the MMPI-A was administered individually due to anticipated reading difficulty. Participants took the MMPI-A on two occasions, once with instructions to underreport and once with standard instructions. The order of the two

¹A copy of the Alcohol and Drug Use Survey (ADUS) is available, upon request, from the first author.

conditions was counterbalanced. Median number of days between testing sessions was 7. Nonsignificant differences on scales between Sessions 1 and 2 within each instructional set (fake good or standard instruction) indicated successful counterbalancing.

The adolescents in the fake-good condition were told to answer the questions to look like they had no problems at home, in school, with friends, or with the law (and that they were functioning well in these areas). They were told to answer in order to look better than they really are and like they have not done the kinds of things young people do to get placed in a detention center or treatment program. We asked them to imagine that they wanted to be released and that to be released, on the test they had to look like they were functioning well and without problems. Finally, we indicated that the person seeing their test scores must not be able to tell that they tried to look better than they really are.

The MMPI-A was usually administered after the ADUS intake. Time between administrations of these tests was not recorded for the purposes of the original study (Stein & Graham, 1999), however, time between tests likely varied from days to weeks (1-14 days).

Compliance check—After all testing was completed, the Compliance Questionnaire was administered to participants.

Scoring and exclusion criteria—Using data provided in the MMPI-A manual (Butcher et al., 1992), for each MMPI-A scale, raw scores were transformed to T scores. According to the MMPI-A manual, protocols with TRIN and VRIN scale T scores greater than or equal to 75 are considered inconsistent (see Butcher et al., 1992). Therefore, those participants with VRIN and TRIN T scores > 74 were excluded from the analyses. Participants with 30 or more unanswered questions were excluded, as were participants with F raw score ≥ 25 (Stein et al., 1995). Exclusions were based on scores produced in both the SI and FG conditions. Participants were excluded if they indicated on the Compliance Questionnaire that they (a) did not comply with the instructions to take the test honestly or the instructions to underreport or (b) did not know on which session they were asked to take the test with standard instructions or on which occasion they were to underreport.

RESULTS

Table 1 presents data for the mixed ANOVA using the within-subjects factor of instructional set and the between-subjects factor of substance abuse group. Main effects and the interaction are presented. Main effects were found on ACK and PRO for both the between (substance abuse group) and within (instructional set) factors. On scale L, only a main effect for the within-factor (instructional set) was found. The interaction (Substance Group $\times$ Instructional Set) was significant only for ACK.

Follow-up tests indicate the following for ACK: Within the FG condition, substance abusers are no different from non-substance abusers on ACK, $F(1, 124) = 1.62, p \geq .10$. Within the SI condition, substance abusers are significantly higher than non-substance abusers on this scale, $F(1, 124) = 30.4, p \leq .001$. Within both the substance-abusing and non-substance-abusing groups, ACK is significantly lower in the FG condition as compared to the SI condition, $F(1, 124) = 133.8, p \leq .001$, for substance abusers; $F(1, 124) = 35.9, p \leq .001$, for non-substance abusers.

Follow-up tests for PRO are similar: Within the FG condition, substance abusers are no different from non-substance abusers, $F(1, 124) = 0.6, p \geq .10$. Within the SI condition, substance abusers are significantly higher than non-substance abusers, $F(1, 124) = 10.1, p$

$\leq .005$. Within both the substance-abusing and non-substance-abusing groups, PRO is significantly lower in the FG condition as compared to the SI condition, $F(1, 124) = 138.5$, $p \leq .001$, for substance abusers; $F(1, 124) = 73.3$, $p \leq .001$, for non-substance abusers.

Follow-up tests indicate the following for L: Within the FG condition, substance abusers are no different from non-substance abusers, $F(1, 124) = 0.6$, $p \geq .10$. Within the SI condition, non-substance abusers are significantly higher than substance abusers, $F(1, 124) = 5.6$, $p \leq .05$. Within both substance- and non-substance-abusing groups, L is significantly higher in the FG condition as compared to the SI condition, $F(1, 124) = 87.0$, $p \leq .001$, for substance abusers; $F(1, 124) = 65.2$, $p \leq .001$, for non-substance abusers.

Comparisons were made between substance abusers in the FG condition and non-substance abusers in the SI condition on scales ACK, PRO, and L. For ACK, $F(1, 124) = 17.9$, $p \leq .005$; for PRO, $F(1, 124) = 53.6$, $p \leq .001$; and for L, $F(1, 124) = 39.8$, $p \leq .001$. Similarly, comparisons were made between substance abusers with SI and non-substance abusers with FG instructions on ACK, PRO, and L. For ACK, $F(1, 124) = 134.6$, $p \leq .001$; for PRO, $F(1, 124) = 129.11$, $p \leq .001$; and for L, $F(1, 124) = 109.27$, $p < .001$.

Table 2 shows classification rates for various groups of substance-abusing and faking teens. This table presents rates for the full sample ($N = 126$) and for a refined sample ($N = 87$) based on ADUS exclusions. Adolescents at the upper end of the "some use" category on the ADUS (scores from 22-50) and adolescents at the lower end of the "abuse" category (ADUS scores from 51-83) may be relatively indistinguishable. Therefore, we analyzed the bottom and top thirds of adolescent scores on the ADUS (scores ≤ 43 vs. scores ≥ 61). This resulted in a sample of $N = 87$.

As shown in Scenario 1 of Table 2, the utility of the substance abuse scales in classifying SI profiles of substance-abusing and non-substance-abusing teens was determined by applying a T-score cutoff of ≥ 55 on ACK and ≥ 60 on PRO (Stein & Graham, 2001). Both ACK and PRO differentiate substance abusers from non-substance abusers, however, ACK performed somewhat better (for $N = 87$, for ACK SN = 0.85, and SP = 0.70; for $N = 126$, ACK SN = 0.67, and SP = 0.69).

As shown in Scenario 2 of Table 2, the utility of the substance abuse scales in classifying FG profiles of substance-abusing and non-substance-abusing teens was determined by applying a T-score cutoff of ≥ 55 on ACK and ≥ 60 on PRO (Stein & Graham, 2001). ACK performed somewhat better than PRO. Contrary to the results found for the SI condition, ACK and PRO did not perform as well in correctly classifying participants. For example, the overall hit rate (OHR) on ACK dropped from .77 (SI condition) to .60 (FG condition) for $N = 87$, whereas OHR on PRO dropped from .63 to .46 for $N = 126$.

Although not the thrust of this article, the relative utility of scale L in differentiating FG and SI condition profiles of the two substance abuse samples was evaluated using the L scale score obtained for all 87 (and 126) profiles. As seen in Scenario 3 of Table 2, an L scale T score ≥ 56 was used to indicate faking good (Stein & Graham, 1999). Table 2 shows that for $N = 87$, SN = 0.78 and SP = 0.75. For $N = 126$, SN = 0.76 and SP = 0.75.

We also combined Scenarios 1 and 2 and compared classification rates for substance abusers taking the test under standard and fake-good instructions (SASI and SAFG, respectively). $BR_{SAFG} = 0.50$ for this comparison. On ACK for $N = 126$, classification rates are as follows: SN = 0.08, SP = .58, PPP = 0.71, NPP = .74, OHR = 0.33. On PRO, SN = 0.02, SP = .61, PPP = 0.33, NPP = .73, and OHR = 0.31. On ACK for $N = 87$, classification rates are as follows: SN = 0.10, SP = .73, PPP = 0.80, NPP = .74, and OHR = 0.41. On PRO, SN = 0.03, SP = .65, PPP = 0.50, NPP = .74, and OHR = 0.34.

Similarly, we combined Scenarios 1 and 2 and compared SAFG versus non-substance abusers with standard instructions (NSASI). On ACK for $N = 126$ ($BR_{SAFG} = 0.53$), classification rates are as follows: SN = 0.46, SP = .47, PPP = 0.86, NPP = .85, and OHR = 0.47. On PRO, SN = 0.46, SP = .41, PPP = 0.78, NPP = .83, and OHR = 0.44. On ACK for $N = 87$ ($BR_{SAFG} = 0.46$), classification rates are as follows: SN = 0.60, SP = .47, PPP = 0.86, NPP = .96, and OHR = 0.53. On PRO, SN = 0.53, SP = .40, PPP = 0.78, NPP = .86, and OHR = 0.46.

Using the following procedures for $N = 87$, we examined the overall utility of the MMPI-A in identifying the profiles of substance abusers who wish to avoid detection: The substance abusers' FG profiles ($n = 40$) and the non-substance abusers' SI profiles ($n = 47$) were first classified using scale L. Following this, the profiles ruled as valid were classified as substance abusing or non-substance abusing using ACK (found to be most accurate across both SI and FG conditions). The L scale correctly classified 29 of 40 (72.5%) substance abusers faking good and 32 of 47 (68.1%) non-substance abusers with standard instructions. Of the 11 profiles of substance abusers faking good that were ruled as valid by scale L, 7 were then classified by ACK as profiles created by non-substance abusers. Thus, only 7 out of 40 (17.5%) of the substance abusers faking good were able to evade detection by both the L scale and ACK. Among the 32 non-substance abusers with standard instructions who were also classified by L as valid, 10 (31.3%) were incorrectly classified as substance abusers by ACK. Of the 11 underreporting substance abusers, ACK accurately detected 4 of 40 teens as substance abusing (or 4 of 11 teens after eliminating using scale L). This reflects a 10% (or 36%) detection rate.

Results were similar for $N = 126$. The substance abusers' FG profiles ($n = 67$) and the non-substance abusers' SI profiles ($n = 59$) were classified using scale L. Next, the profiles ruled as valid were classified as substance abusing or non-substance abusing using ACK. The L scale correctly classified 48 of 67 (71.6%) substance abusers faking good and 41 of 59 (69.5%) non-substance abusers with standard instructions. Of the 19 profiles of substance abusers faking good that were ruled as valid by scale L, 13 were then classified by ACK as profiles created by non-substance abusers. Therefore, only 13 out of 67 (19.4%) of the substance abusers faking good were able to evade detection by both the L scale and ACK. Among the 41 non-substance abusers with standard instructions who were also classified by L as valid, 13 (31.2%) were incorrectly classified as substance abusers by ACK. Of the 19 under-reporting substance abusers, ACK accurately detected 6 of 67 teens as substance abusing (or 6 of 19 teens after eliminating using scale L). This reflects a 9% (or 32%) detection rate.

DISCUSSION

Incarcerated substance abusers and incarcerated non-substance abusers can significantly diminish scores on ACK and PRO when instructed to conceal problems and symptoms. As expected, under SI conditions, substance abusers produced higher scores than non-substance abusers on ACK and PRO and lower scores on L. Also as expected, for both substance abusers and non-substance abusers, ACK and PRO are higher and L is lower under SI conditions as compared to FG conditions. Results indicate no differences between substance abusers and non-substance abusers under FG conditions on ACK and PRO. Unfortunately, these scales are not resistant to faking good. However, fortunately, L is not affected by substance abuse status: Under FG conditions, no differences were found between substance abusers and non-substance abusers on L. It is reassuring that substance abuse status does not impact performance on scale L under FG conditions.

Substance abusers faking good produced higher scores on scale L than did non-substance abusers with SI. Similarly, non-substance abusers faking good produce higher scores on scale L than do substance abusers with SI. As expected, substance abusers with SI produced higher ACK and PRO scores than did non-substance abusers faking good. As compared to non-substance abusers with SI, substance abusers faking good produced significantly lower scores on ACK and PRO. Again, these findings illustrate the L scale's ability to distinguish faking from SI conditions and the vulnerability of substance abuse scales to faking.

PRO performed more poorly than ACK in classifying substance abusers faking good; however, ACK also evidenced poor classification rates under the FG condition. Thus, although ACK and PRO had acceptable classification rates when applied to honest profiles (see Scenario 1, Table 2), their accuracy decreased substantially when the substance abusers were motivated to avoid detection (Scenario 2). For example, using $N = 87$, $PRO \geq 60$ produced $SN = 0.08$. On the other hand, non-substance abusers had very good classification rates under FG conditions ($ACK < 55$ and $PRO < 60$ produced $SP = 0.93$ and 0.88 , respectively, at $N = 126$). These results are similar to those presented by Otto et al. (1988), who found relatively high classification rates for substance abuse scales under SI conditions that drastically diminished under FG conditions.

ACK appeared to be the best substance abuse scale under SI and FG conditions. Nonetheless, there was still significant disagreement between its classifications and the criterion as evidenced by the fact that ACK correctly identified only 8 of 40 (Sensitivity $[SN] = .20$) and 11 of 67 ($SN = .16$) substance abusers faking good for $N = 87$ and 126, respectively (see Table 2, Scenario 2). Because Scale L performed well in distinguishing valid and invalid profiles, our findings that substance abusers can successfully avoid detection when motivated to underreport is less troubling. However, it should be acknowledged that the MMPI-A is of limited specific utility in cases where nothing more can be said about a profile except that it is invalid. Otto et al. (1988) found that scale L misclassified 61% of alcoholics faking good, and this is in stark contrast to the current study. The good performance of scale L is consistent with previous research showing that it generally performs as well as or better than other dissimulation measures in adolescents (see Stein & Graham, 1999).

The limitations of ACK and PRO, without first considering scale L, are clearly illustrated by classification rates obtained when comparing SAFG versus SASI groups and SAFG versus NSASI groups. Both comparisons produced very poor SN, SP, and OHR for both $N = 87$ and $N = 126$ ($Mdn = 0.46$). Predictive powers obtained were generally higher than SN and SP because adolescents predicted to meet both conditions (e.g., substance abuse and faking good) are fewer in number than adolescents actually meeting the predicted conditions (and because predictive power is based on numbers of adolescents predicted to meet a condition). ACK generally produced acceptable predictive powers for both comparisons, whereas PRO produced acceptable predictive powers for only the SAFG versus NSASI comparison.

Use of the L scale and ACK together was particularly encouraging. The L scale first correctly classified most of the non-substance abusers under the SI condition as well as the substance abusers faking good. Of the few substance abusers' faked profiles that escaped detection as invalid, only 17.5% were not subsequently classified as substance abusive by ACK for the sample of $N = 87$ (this number is 19.4% for $N = 126$). Although Otto et al. (1988) found a slightly lower rate (about 8%), their results are generally consistent with results presented here. Of non-substance abusers with standard instructions who were classified by L as valid, as many as 31.3% were incorrectly classified as substance abusers by ACK (68.7% correctly classified). Although this rate is somewhat disappointing, using

this optimal cut score on ACK (≥ 55) may be acceptable because for many delinquent adolescents, this is their first opportunity for treatment.

With a misclassification rate as high as 31.3%, it is important to note that testing is only part of an assessment. Assessment typically should consist of direct observation, face-to-face interview, mental status exam, chart review, interview with collaterals (parents), and family history (Wasserman et al., 2003). Clinicians must look for converging information for final determination regarding substance abuse status. Oftentimes, clinicians may have only testing and interview data (Wasserman et al., 2003). In these cases, when testing and interview data conflict, clinicians may consider placing more emphasis on testing results because adolescents may be more honest during computer-driven and paper-pencil testing than in face-to-face interviews (Turner, Lessler, & Devore, 1992; Waterton & Duffy, 1984).

Alternatively, another option for clinicians is to examine item endorsement on ACK. Of 13 items, the content of 6 items does not relate directly to substance use. It is possible that non-substance-abusing teens are misclassified as substance abusing if most or all of these 6 items are endorsed with few of the substance-related items endorsed. Future research may examine this issue more directly and develop a scale in which all items have content-related substance use.

Results indicate clear limitations on the substance abuse scales used in isolation without benefit of scale L. Recall, of 19 underreporting substance abusers, ACK accurately detected only 6 of 67 teens as substance abusing (or 6 of 19 teens after eliminating using scale L). This reflects a 9% (or 32%) detection rate. Similarly, of 11 underreporting substance abusers, ACK accurately detected 4 of 40 teens as substance abusing (or 4 of 11 teens after eliminating using scale L). This reflects a 10% (or 36%) detection rate. As shown in prior research reviewed here, scale L is very useful but nonspecific in terms of identifying what is being minimized. For adolescents producing invalid profiles, further assessment is required, which may then detect substance abuse status.

Results of this study suggest that although substance abusers may be able to decrease elevations on substance use scales or avoid detection by the L scale under FG conditions, it is rare that they can do both. This is consistent with Otto et al. (1988). For incarcerated adolescents, scale L and ACK may be the best combination to use when attempting to identify incarcerated substance abusers using the MMPI-A. Results of this study should be used cautiously because base rates (BR) for substance abuse and faking good may influence results substantially. However, we note that BRs of substance use presented here (BR = 0.53, $N = 126$; BR = 0.46, $N = 87$) are not uncommon in this setting. Replication is recommended in other samples and settings. Similarly, caution is in order because development of multiple cut scores using L first and ACK second likely capitalizes on chance variation.

We tentatively offer suggestions for clinicians. The current data suggest that when the L scale is ≥ 56 , there is very strong possibility that the adolescent is underreporting. No further interpretation of the protocol should take place. For protocols with $L \leq 56$, when $ACK \geq 55$ there is strong possibility that the teen misuses substances. These adolescents should be further assessed to confirm. In the current study, about 30% of non-substance-abusing teens were misclassified as substance abusing. When $ACK \leq 55$, it appears that it is likely the teen does not misuse substances. Recall, in the current study, less than 20% of substance abusers were missed by ACK. More research should accumulate on this matter before such decision rules are considered standard.

The results of this study are consistent with previous research indicating the usefulness of content-based (face-valid) scales such as ACK in comparison with empirically derived

scales such as PRO (see Archer, Elkins, Aiduk, & Griffin 1997; Ben-Porath, McCully, & Almagor, 1993; Stein, Graham, Ben-Porath, & McNulty, 1999). We utilized somewhat general instructions for the FG condition and did not request that teens focus specifically on underreporting substances. It appears that this is consistent with directions provided by Otto et al. (1988). It is important to note, however, that the instructions we utilized cover a variety of problems, including, but not limited to, substance abuse difficulties. In addition, such general instructions are more ecologically valid in this setting because teens may underreport a host of behaviors and attitudes during testing. At the same time, it remains possible that ACK is more susceptible to distortion of substance use specifically as compared to empirically keyed scales such as PRO. Furthermore, we do not know the impact on scale L of instructions to only underreport substance use. Scale L may not pick up adolescents faking in only a particular area. Future studies may address susceptibility of L and of ACK compared to PRO when respondent defensiveness is specific to substance use. On a related note, future studies may examine the impact of coaching on dissimulated substance use. There is some evidence that coaching on how to dissimulate may reduce detection of dissimulators (see Rogers & Kelly, 1997). Finally, we recommend that future studies increase *N* to examine characteristics of adolescent substance abusers who are able to successfully fake-good.

A limitation of this study is that both interrater reliability and temporal stability data were unavailable for substance abuse ratings using the ADUS. However, it is important to note that this instrument was not introduced to this setting for the purposes of this study. On the contrary, the ADUS was an integral part of screening and assessment procedures for adolescents entering the facility. In this context, it was impractical for us to ask that the ADUS be administered to a subsample of participants on two occasions by the counselors.

Although no psychometric data have been published on the ADUS, it may have some advantages over diagnostic data found in a teen's records. First, it is unlikely that a structured and standardized interview was used in determining substance use diagnoses found in records. According to Pollock, Martin, and Langenbucher (2000), there is disagreement in the conceptual framework and definition of alcohol abuse disorder across nosological systems such as DSM-IV and ICD-10, even when using a structured interview. Second, we also did not know the qualifications of persons who assigned substance use diagnoses. On the other hand, the ADUS was a structured interview provided by well-trained CCDCs. Third, the ADUS was usually administered only days or weeks prior to the MMPI-A. This temporal relationship is clinically meaningful because the MMPI-A would be used to help determine an adolescent's needs at the time of assessment. Fourth, the clinical judgment used in assigning substance use diagnoses is often based on adult models, which may not apply to adolescents (Winters, 1990). On the other hand, the ADUS was derived specifically for adolescents and CCDCs were trained specifically to work with juvenile offenders.

Identifying potential substance-abuse problems in this setting can be critical for understanding factors that may have been involved in misconduct as well as treatment placement. There may be occasions when an adolescent is not honest during testing, and this argues for the usefulness of tests with accurate faking scales. Testing can alert professionals to areas needing further assessment and can assist in streamlining interviews. This study indicates that the MMPI-A can play an important role in screening and assessing for substance abusers attempting to underreport in juvenile correctional settings. The unique contribution of this study lies in its examination of how well the MMPI-A detects substance abusers who fake. More research is needed on this important issue.

Acknowledgments

This work was supported by a grant from the National Institute on Drug Abuse (R01, #13375; PI, Stein). The authors gratefully acknowledge Beverly Kaemmer and the University of Minnesota Press for their support in completing this project. We thank Suzanne Sales and Cheryl Eaton for their invaluable technical assistance in this work.

Biography

L. A. R. Stein is assistant professor of research at the Center for Alcohol and Addiction Studies of Brown University. She received her Ph.D. from Kent State University. She is currently principle investigator and co-investigator on several National Institute on Drug Abuse (NIDA)–funded studies examining incarcerated persons and substance use issues. She also directs a postdoctoral clinical program in juvenile forensic psychology and is the director of research for the Rhode Island Training School.

John R. Graham is professor of psychology at Kent State University. He received his Ph.D. at the University of North Carolina at Chapel Hill. He has authored and coauthored a myriad of books and peer-reviewed articles on assessment of adults and adolescents. His primary interests are in personality and forensic assessment using the MMPI-2 and MMPI-A. Specific interests include (a) identification of invalid responding, (b) validity with ethnic minority groups, (c) content and content-component scales, (d) correlates of scales and code types in a variety of settings, and (e) predicting institutional and postinstitutional adjustment of prisoners. He has been editor and on the editorial boards for several professional journals. He has held positions of department chair and clinical director in the Department of Psychology, Kent State University.

REFERENCES

- Archer, RP. MMPI-A: Assessing adolescent psychopathology. 2nd ed.. Lawrence Erlbaum; Hillsdale, NJ: 1997.
- Archer RP, Elkins DE, Aiduk R, Griffin R. The incremental validity of MMPI-2 supplementary scales. *Assessment* 1997;1:193–205.
- Baer RA, Ballenger J, Kroll LS. Detection of under-reporting on the MMPI-A in clinical and community samples. *Journal of Personality Assessment* 1998;71:98–113.
- Basham RB. Clinical utility of the MMPI research scales in the assessment of adolescent acting out behaviors. *Psychological Assessment* 1992;4:483–492.
- Ben-Porath YS, McCully E, Almagor M. Incremental validity of the MMPI-2 content scales in the assessment of personality and psychopathology by self-report. *Journal of Personality Assessment* 1993;61:193–205.
- Butcher, JN.; Dahlstrom, WG.; Graham, JR.; Tellegen, A.; Kaemmer, B. MMPI-2: Manual for administration and scoring. University of Minnesota Press; Minneapolis: 1989.
- Butcher, JN.; Han, K. Development of an MMPI-2 scale to assess the presentation of self in a superlative manner: The S Scale.. In: Butcher, JN.; Spielberger, CD., editors. *Advances in personality assessment*. Vol. 10. Lawrence Erlbaum; Hillsdale, NJ: 1995. p. 25-50.
- Butcher, JN.; Williams, CL.; Graham, JR.; Archer, RP.; Tellegen, A.; Ben-Porath, YS., et al. MMPI-A: Manual for administration, scoring, and interpretation. University of Minnesota Press; Minneapolis: 1992.
- Cashel ML, Rogers R, Sewell KW, Holliman NB. Preliminary validation of the MMPI-A for a male delinquent sample: An investigation of clinical correlates and discriminant validity. *Journal of Personality Assessment* 1998;71:49–69. [PubMed: 9807230]
- Cofer CN, Chance J, Judson AJ. A study of malingering on the MMPI. *Journal of Psychology* 1949;27:491–499. [PubMed: 18111588]

- Edwards, AL. The social desirability variable in personality assessment and research. Dryden; New York: 1957.
- Gallucci NT. Correlates of MMPI-A substance abuse scales. *Assessment* 1997;4:87–94.
- Gantner AB, Graham JR, Archer RP. Usefulness of the MAC scale in differentiating adolescents in normal, psychiatric, and substance abuse settings. *Psychological Assessment* 1992;4:133–137.
- Gottesman II, Prescott CA. Abuses of the MacAndrew MMPI alcoholism scale: A critical review. *Clinical Psychology Review* 1989;9:223–242.
- Hathaway SR, McKinley JC. A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology* 1940;10:249–254.
- Hawkins, CE. Mental health survey. Sep. 1987 Unpublished interoffice communication on the Ohio Mental Health Prevalence Study..
- Herkov MJ, Archer RP, Gordon RA. MMPI response sets among adolescents: An evaluation of the limitations of the subtle-obvious subscales. *Psychological Assessment: A Journal of Consulting and Clinical Psychology* 1991;3:424–426.
- MacAndrew C. The differentiation of male alcoholic outpatients from nonalcoholic psychiatric outpatients by means of the MMPI. *Quarterly Journal of Studies on Alcohol* 1965;26:238–246. [PubMed: 14320345]
- McNulty JL, Graham JR, Ben-Porath YS, Stein LAR. A comparative study of the validity of MMPI-2 scores in African American and Caucasian outpatients. *Psychological Assessment* 1997;9:464–470.
- Miller, GA. The SASSI Adolescent manual. SASSI Institute; Bloomington, IN: 1990.
- Otto RK, Lang AR, Megargee EI, Rosenblatt AI. Ability of alcoholics to escape detection by the MMPI. *Journal of Consulting and Clinical Psychology* 1988;56:452–457. [PubMed: 3397437]
- Pena LM, Megargee EI, Brody E. MMPI-A patterns of male juvenile delinquents. *Psychological Assessment* 1996;8:388–397.
- Pollock NK, Martin CS, Langenbucher JW. Diagnostic concordance of DSM-III, DSM-III-R, DSM-IV and ICD-10 alcohol diagnoses in adolescents. *Journal of Studies on Alcohol* 2000;61:439–446. [PubMed: 10807216]
- Rogers R, Cashel ML, Johansen J, Sewell KW, Gonzalez C. Evaluation of adolescent offenders with substance abuse: Validation of the SASSI with conduct-disordered youth. *Criminal Justice and Behavior* 1997;24:114–128.
- Rogers, R.; Kelly, KS. Denial and misreporting of substance abuse.. In: Rogers, R., editor. *Clinical assessment of malingering and deception*. Guilford; New York: 1997. p. 108-129.
- Stein LAR, Colby SM, O'Leary T, Monti PM, Rohsenow D, Spirito A, et al. Response distortion in adolescents who smoke: A pilot study. *Journal of Drug Education* 2002;32:271–286. [PubMed: 12556133]
- Stein LAR, Graham JR. Detecting fake-good MMPI-A profiles in a correctional facility. *Psychological Assessment* 1999;11:386–395.
- Stein LAR, Graham JR. Use of the MMPI-A to detect substance abuse in a juvenile correctional setting. *Journal of Personality Assessment* 2001;77:508–523. [PubMed: 11781036]
- Stein LAR, Graham JR, Ben-Porath YS, McNulty JL. Using the MMPI-2 to detect substance abuse in an outpatient mental health setting. *Psychological Assessment* 1999;11:94–100.
- Stein, LAR.; McClinton, BK.; Graham, JR. Examining possible ethnic differences on the MMPI-A.. Paper presented at the 30th annual Symposium on Recent Developments in the Use of the MMPI (MMPI-2 and MMPI-A); Tampa, FL. Mar. 1995
- Sweeney, B. Alcohol and Drug Use Survey (ADUS). 1987. Unpublished survey instrument for the Ohio Dept. of Youth Services. Funded by the Ohio Governor's Office of Criminal Justice Services in cooperation with the Ohio Dept. of Mental Health..
- Teplin LA, Abram KM, McClelland GM, Dulcan MK, Mericle AA. Psychiatric disorders in youth in juvenile detention. *Archives of General Psychiatry* 2002;59:1133–1143. [PubMed: 12470130]
- Turner, C.; Lessler, J.; Devore, J. The effects of mode of administration and wording on reporting of drug use.. In: Turner, CF.; Lessler, JT.; Gfroerer, JC., editors. *Survey measurement of drug use*:

- Methodological studies. U.S. Dept. of Health & Human Services, Alcohol, Drug Abuse, and Mental Health Administration; Rockville, MD: 1992.
- Wasserman G, Jensen P, Ko S, Coccoza J, Trupin E, Angold A, et al. Mental health assessments in juvenile justice: Report on the Consensus Conference. *Journal of the American Academy of Child and Adolescent Psychiatry* 2003;42:752–761. [PubMed: 12819434]
- Waterton JJ, Duffy JC. A comparison of computer interviewing techniques and traditional methods in the collection of self-report alcohol consumption data in a field survey. *International Statistical Review* 1984;52:173–182.
- Weed NC, Butcher JN, Williams CL. Development of MMPI-A alcohol/drug problem scales. *Journal of Studies on Alcohol* 1994;55:296–302. [PubMed: 8022177]
- Wiggins JS. Interrelationships among MMPI measures of dis-simulation under standard and social desirability instructions. *Journal of Consulting Psychology* 1959;23:419–427.
- Winters K. The need for improved assessment of adolescent substance involvement. *The Journal of Drug Issues* 1990;20:487–502.
- Winters KC. A new multiscale measure of adult substance abuse. *Journal of Substance Abuse Treatment* 1999;16(3):237–246. [PubMed: 10194741]
- Wolfson KP, Erbaugh SE. Adolescent responses to the MacAndrew Alcoholism Scale. *Journal of Consulting and Clinical Psychology* 1984;52:625–630. [PubMed: 6470288]

TABLE 1

Mixed ANOVA of Substance Group (Between Factor: Substance Abuse and Non-Substance Abuse) and Instruction Set (Within Factor: Faking Good and Standard Instruction)

Scale	Substance Abuse		Non-Substance Abuse		Main Effects: Substance Group		Main Effects: Instruction Set		Interaction: Substance $\times$ Instruction	
	Fake Good	Standard Instruction	Fake Good	Standard Instruction	F (1, 124)	η^2	F (1, 124)	η^2	F (1, 124)	η^2
ACK										
M	43.6	62.3	41.3	51.6	20.5***	0.14 ^a	150.9***	0.55 ^a	12.6***	0.09 ^b
SD	11.3	11.7	8.0	9.7						
PRO										
M	41.4	63.0	39.8	56.5	6.6**	0.05 ^b	204.4***	0.62 ^a	3.3*	0.03 ^c
SD	11.6	11.2	11.7	11.5						
L										
M	68.9	48.5	71.2	52.4	2.9*	0.02 ^c	150.6***	0.55 ^a	0.3*	.00
SD	17.8	8.5	15.3	9.9						

NOTE: ACK = Addiction Acknowledgement Scale; PRO = Addiction Process Scale; L = Lie Scale. Scales are in linear T scores.

^aEffect size (ES) = large.

^bES = medium.

^cES = small.

*
 $p \geq .10$.

**
 $p \leq .05$.

 $p \leq .001$.

TABLE 2

Sensitivity (SN), Specificity (SP), Positive Predictive Power (PPP), Negative Predictive Power (NPP), and Overall Hit Rates (OHR) Under Three Classification Scenarios

Scale Score	N = 126						N = 87					
	SN	SP	PPP	NPP	OHR		SN	SP	PPP	NPP	OHR	
1. Substance abusers and non-substance abusers under standard instructions												
ACK ≥ 55	.67	.69	.71	.65	.68		.85	.70	.71	.85	.77	
PRO ≥ 60	.67	.58	.64	.61	.63		.73	.62	.62	.73	.67	
2. Substance abusers and non-substance abusers faking good												
ACK ≥ 55	.16	.93	.73	.50	.52		.20	.94	.73	.58	.60	
PRO ≥ 60	.09	.88	.46	.46	.46		.08	.91	.43	.54	.53	
3. Faking good and standard instruction conditions												
L ≥ 56	.76	.75	.76	.75	.76		.78	.75	.76	.77	.76	

NOTE: ACK = Addiction Acknowledgement Scale, PRO = Lie Scale. Scales are in linear T scores. SN = teens with target condition correctly classified. Target conditions: substance abuse (Scenarios 1 & 2), faking good (Scenario 3); SP = teens without target condition correctly classified; PPP = probability that a teen identified by a scale actually meets the target condition; NPP = probability that a teen not identified by a scale actually does not meet the target condition; OHR = all teens correctly classified in target and nontarget conditions. Base Rates (BR) for Scenario 1 with N = 126, Scenario 2 with N = 87, and Scenario 3 with N = 126 and Scenario 3 with N = 87 is .50.