EXTERNAL VALIDATION OF THE MMPI-A-RF WITH YOUTH WITH MENTAL HEALTH NEEDS: A SYSTEMATIC EXAMINATION OF SYMPTOM-BASED CORRELATES AND INTERPRETIVE STATEMENTS

Allyson J. Sharf, M.S.

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

August 2019

APPROVED:

Richard Rogers, Major Professor
Amy Murrell, Committee Member
Randall Cox, Committee Member
Vicki Campbell, Chair of the Department of Psychology
David Holdeman, Dean of the College of Liberal Arts and Social Sciences
Victor Prybutok, Dean of the Toulouse Graduate School

Sharf, Allyson J. *External Validation of the MMPI-A-RF with Youth with Mental Health Needs: A Systematic Examination of Symptom-Based Correlates and Interpretive Statements*. Doctor of Philosophy (Clinical Psychology), August 2019, 145 pp., 25 tables, 1 appendix, references, 129 titles.

Over the last several decades, the Minnesota Multiphasic Personality Inventory - Adolescent (MMPI-A) has remained one of the most extensively studied and commonly used adolescent assessment measures. Most recently, the MMPI-A was revised, published as the Minnesota Multiphasic Personality Inventory - Adolescent Restructured Form (MMPI-A-RF). Given the infancy of the extant MMPI-A-RF literature, the current dissertation sought to be one of the first criterion studies since the test manual to establish its relevant clinical correlates in a sample of youth referred for mental health services. Following in line with previous clinical correlate research, the psychometric properties of the MMPI-A-RF were determined. Under this approach, the present study bolstered support for the construct validity of the RC scales through the identification of clinically relevant, scale and item-level correlates. In particular, the RCd and RC4 evidenced especially strong convergent and discriminant validity. As a result, the current study highlighted the MMPI-A-RF's efficacy in detecting psychopathology along the broad externalizing and internalizing spectrum. In addition, the role of gender and its effect on MMPI-A-RF clinical interpretations were explored. Practical recommendations for ensuring a gender-specific approach to MMPI-A-RF interpretation were outlined.

TABLE OF CONTENTS

Page

LIST OF TABLES v

CHAPTER 1. INTRODUCTION 1

Overview of Adolescent Measures 3

Development of the MMPI 9

MMPI with Adolescents 10

MMPI-A 18

MMPI-A-RF 23

Methodological Considerations in Validating Clinical Correlates 24

Research Methodology 25

Gendered vs Non-Gendered Norms 29

Current Study 31

Research Questions and Hypotheses 33

Supplementary Research Questions 34

CHAPTER 2. METHOD 35

Design 35

Operationalization of Construct Validity 35

Participants 37

Study Measures 37

Kiddie Schedule for Affective Disorders and Schizophrenia-Present and Lifetime Version (K-SADS-PL) 37

Massachusetts Youth Screening Instrument - Version 2 (MAYSI-2) 38

Minnesota Multiphasic Personality Inventory - Adolescent Restructured Form (MMPI-A-RF) 39

Wide Range Achievement Test – Fourth Edition (WRAT-4) – Word Reading Subtest 39

Procedure 40

Recruitment and Informed Consent 40

Standardized Measures 40

Manipulation Check and Debriefing 41

CHAPTER 3. RESULTS 42
Refinement of the Sample 42
Description of the Final Sample 42
Research Questions and Hypotheses 47
Research Question 1 47
Research Question 2 49
Research Question 3? 55
Research Question 4 58
Research Question 5 75
Research Question 6 78
Supplementary Analyses 86
Supplementary Research Question 1 86
Supplementary Research Question 2 89
Supplementary Research Question 3 91
CHAPTER 4. DISCUSSION 93
The Clinical Utility of the MMPI-A-RF 96
Scale Homogeneity of the MMPI-A-RF Reconstructed Clinical Scales 96
Construct Validity of the MMPI-A-RF Reconstructed Clinical Scales 98
Examination of Interpretive Statements for Reconstructed Clinical Scales 101
Differences between Higher Order and Reconstructed Clinical Scales 109
The Importance of Gender for Interpretive Statements 113
Clinical Implications 119
Top-Down Approach to MMPI-A-RF Empirical Correlates 119
Strengths and Limitations of Empirically Based Interpretations for MMPI-A-RF RC Scales 122
Gender Differences and MMPI-A-RF Empirical Correlates 123
Limitations and Future Directions 124
Limitations 124
Future Directions 125
Final Considerations 126
APPENDIX: SUPPLEMENTARY TABLES 128
REFERENCES 136

LIST OF TABLES

Page

Table 1. Characteristics of the MMPI-A-RF Clinical Correlate Studies 26

Table 2. Gender, Ethnicity, and Language Differences between Partial Hospitalization and Juvenile Detention Settings 44

Table 3. Demographic Differences between Partial Hospitalization and Juvenile Detention Settings 45

Table 4. Differences in Mean Symptom Severity by Common Adolescent Disorder between Partial Hospitalization and Juvenile Detention Settings 46

Table 5. Alphas and Mean Inter-item Correlations of MMPI-A-RF RC Scales 48

Table 6. Percentage Evidencing Scaling Success and Cross-Scale Correlations for the MMPI-A-RF RC Scales 49

Table 7. Substantial Correlations (> .35) of MMPI-A-RF RC Scales with K-SADS-PL Diagnostic Section Scores 51

Table 8. Substantial Correlations (> .35) of MMPI-A-RF Externalizing Subscales with K-SADS-PL Externalizing Diagnostic Section Scores 54

Table 9. Effectiveness of MMPI-A-RF RC7 Cut Scores (RC7 > 60) for Identifying Major Depressive Disorder and Generalized Anxiety Disorder Diagnoses 56

Table 10. Effectiveness of MMPI-A-RF RC4 Cut Scores for Identifying Conduct Disorder and Substance Use Disorder Diagnoses 58

Table 11. Substantial K-SADS-PL Internalizing Item-Level Correlates (> .35) for the MMPI-A-RF RC Scales 59

Table 12. Substantial K-SADS-PL PTSD Item-Level Correlates (> .35) for the MMPI-A-RF RC Scales in a Sample of Adolescents Meeting Criteria for PTSD 66

Table 13. Substantial K-SADS-PL Externalizing Item-Level Correlates (> .35) for the MMPI-A-RF RC Scales 68

Table 14. Substantial K-SADS-PL Externalizing Item-Level Correlates (> .35) for the MMPI-A-RF Externalizing Subscales 72

Table 15. Differences in MMPI-A-RF RC Scale Elevations between Females and Males 76

Table 16. Substantial Correlations (> .35) of MMPI-A-RF RC Scales by Gender with the K-SADS-PL Diagnostic Section Total Scores 77

Table 17. Substantial Correlations (> .35) of MMPI-A-RF H-O scales with K-SADS-PL Diagnostic Section Scores 79

Table 18. Substantial K-SADS-PL Internalizing Item-Level Correlates (> .35) for the MMPI-A-RF H-O Scales 81

Table 19. Substantial K-SADS-PL Externalizing Item-Level Correlates (> .35) for the MMPI-A-RF H-O Scales 83

Table 20. Differences in Relevant K-SADS-PL Symptomatology Total Scores by Caution/Warning Range Classifications for MAYSI-2 Scales 87

Table 21. Effectiveness of MAYSI-2 Caution/Warning Classifications for Suicidal Ideation Scale 88

Table 22. Preliminary Analysis of Substantial Correlations (> .35) of MMPI-A-RF RC scales with the K-SADS-PL Total Scores for WNL Profiles (n = 9) 90

Table 23. MMPI-A-RF RC4 and AGG-r Scales as Predictors of Aggressive Conduct Problems 91

Table 24. Comparison of Correlations between RC7 and Selected Depressive/Anxious Symptomatology in MMPI-A-RF Clinical Correlate Studies 108

Table 25. Gender Differences between RC7 and RC8 Correlations with PTSD Symptomatology in MMPI-A-RF Clinical Correlate Studies 116

CHAPTER 1

INTRODUCTION

Youth frequently suffer from severe mental health problems, some known and others undiagnosed. In fact, research (Costello, Mustillo, Erkanli, Keeler, & Angold, 2003) estimates approximately 35% of sixteen year olds meet criteria for at least one mental disorder. For those exhibiting significant mental health issues, referrals to mental health services are often made. Partial hospitalization programs represent an increasingly popular option for more intensive mental health treatment (Thatt, Makinen, Nguyen, Hill, & Flament, 2013).

For some youth, the juvenile justice system constitutes their referral into receiving mental health services. As with adults, a trend has emerged in which the juvenile justice system is increasingly housing and treating the mentally ill (Cocozza & Skowyra, 2000). Research demonstrates that more than 70% of legally involved youth met criteria for at least one psychiatric diagnosis (Odgers, Burnette, Chauhan, Moretti, & Reppucci, 2005). This prevalence rate is approximately double than that of the general population (Costello et al., 2003). Such high prevalence rates underscore the increased mental health concerns in adolescents as well as the need for their accurate assessment.

Particularly for youth, early identification of psychological symptoms is essential for timely interventions and positive treatment outcomes. If psychopathology is left untreated, juveniles are at much greater risk for serious adverse consequences such as increased likelihood of suicide, substance dependence, or criminal recidivism (Fazel, Doll, & Langstrom, 2008; Russel & Marston, 2010). Therefore, it is critical for mental health professionals to have an array of effective assessment tools at their disposal that can accurately evaluate disordered juveniles suffering from psychological difficulties.

This dissertation addresses the need for rigorously validated, adolescent-specific multiscale inventories. For adults, the Minnesota Multiphasic Personality Inventory-2nd Edition (MMPI-2; Butcher et al., 2001) and Minnesota Multiphasic Personality Inventory-2nd Edition Restructured Form (MMPI-2-RF; Ben-Porath & Tellegan, 2008) represent the most widely used and well-validated measures to assess patterns of psychopathology. However, there remains an important need to develop and validate multiscale inventories specifically for adolescents. Consequently, the MMPI-Adolescent-Restructured Form (MMPI-A-RF; Archer, Handel, Ben-Porath, & Tellegan, 2016) was recently developed and published. As such, the MMPI-A-RF remains largely untested in relevant adolescent populations, such as partially hospitalized youth or juvenile offenders. The current dissertation seeks to investigate the clinical utility of the MMPI-A-RF with these populations using a well-validated semi-structured interview. Semi-structured interviews hold several advantages over other assessment methods in terms of reliability and validity, thus enhancing the current dissertation's methodological approach.

This chapter is organized into four major sections that address various issues surrounding adolescent assessments. The first section provides context for the current dissertation through a brief overview of the methods employed in adolescent assessments, including interviews and multiscale inventories. The second section builds on the previous section by detailing the development and evolution of the most commonly used multiscale inventories, the MMPI family. It also serves to introduce the principal measure of the current study—the MMPI-A-RF. Then, the third section reviews the different considerations in validating symptom-based correlates and clinical descriptors, a critical interpretive issue facing all new measures such as the MMPI-A-RF. The fourth and final section introduces the current study. It outlines the

research questions and hypotheses designed to validate the clinical interpretations of the MMPI-A-RF.

Overview of Adolescent Measures

Clinical assessment has always constituted an essential aspect of the practice of professional psychology; however, the specific methods employed have continually changed over time. Beginning with the traditional (i.e., unstructured) clinical interviews initially used by psychologists decades ago, this section outlines the evolution of assessment measures used to evaluate psychopathology in adolescents.

Historically, unstructured clinical interviews are considered to be the original form of psychological assessment (Groth-Marnat, 2009). In general, traditional clinical interviews sought to develop a psychological portrait of the client, including the person's current difficulties and any diagnoses. Its time-honored format relies on an open-ended question and answer format without requiring any preset structure regarding the order of topics considered (Craig, 2003; Groth-Marnat, 2009). Given its flexibility, it remains the sole responsibility of the clinician in "deciding what questions to ask and how the resulting information is used in arriving at a diagnosis" (Summerfeldt & Antony, 2002, p. 3).

The use of traditional unstructured clinical interviews has continued to the present day (Craig, 2003; Jones, 2010), clearly due to several advantages. Mainly, unstructured clinical interviews allow for great flexibility on the part of the interviewer in terms of the focus, depth, or style of an interview (Craig, 2003). Thus, an adept clinician can easily customize an interview to meet the individual needs of each client. For example, Shedler (2002) purposefully integrated such flexibility into the administration of the Shedler and Westen Assessment Procedure-200

(Westen & Shedler, 1999). As Shedler (2002) explained, the administration of the SWAP-200 relies on "the empathically attuned and dynamically sophisticated clinician given free rein to practice his or her craft" (p. 433).

Increased flexibility also benefits the clients in terms of their feelings of control over the interview. In an unstructured approach, clients are able to discuss their experiences in more of a story-telling format (Moyle, 2001). As such, clients provide input concerning the direction of the interview by introducing relevant information where it most makes sense to them. In this way, interviewers do not unduly influence the client's responses (Moyle, 2001). Moreover, allowing clients to have more influence on the interview can be especially helpful in establishing rapport (Turner, Hersen, & Heiser, 2003).

Despite their strengths, the unstructured nature of the interviews has been criticized for several practical reasons. First, unstructured clinical interviews are intrinsically associated with a wide degree of variability. In their seminal work, Ward, Beck, Mendelson, Mock, and Erbaugh (1962) documented the major sources of such variability as applied to diagnosis. Ward and colleagues (1962) concluded that many more diagnostic disagreements stemmed from *method* variability (i.e., how the interview was conducted) as opposed to *patient* variability (i.e., differences in his or her clinical presentation). More specifically, nearly one-third (32.5%) of the variability resulted from interview styles. Another 62.5% of the differences stemmed from clinicians' differences in the criterion for what constituted a salient symptom. In contrast, variability in patient presentation tended to be a minimal influence (i.e., 5%). Ward et al. (1962) underscored the notion that how unstructured interviews are conducted can strongly influence clinicians' conclusions.

One particularly troubling result of the unstructured approach to interviews is the potential for missed diagnoses or misdiagnosis (Rogers, 2003). As a reference, "missed diagnosis" denotes not diagnosing a mental disorder when it is present, whereas "misdiagnosis" refers to diagnosing a mental disorder that is not present. Empirical research has consistently documented wide variability in terms of agreement between unstructured clinical interviews and their more structured counterparts (Craig, 2003; North et al., 1997; Rettew, Lynch, Achenbach, Dumenci, & Ivanova, 2009; Rogers, 2003). In a meta-analysis of 38 studies, Rettew and colleagues (2009) found very low to moderate agreement (mean kappas ranged from 0.19 to 0.64) between assessment methods for most clinical diagnoses. When examined more closely, unstructured clinical interviews tended to under-diagnose mental disorders as compared to structured interviews. As one example, major depressive disorder prevalence rates were higher when employing structured interviews (i.e., 26%) than studies taking an unstructured approach (i.e., 17%).

According to Rettew and colleagues (2009), several aspects of unstructured interviews limit agreement with more structured interviews and ultimately lead to under-diagnosis. Whereas structured approaches require the interviewer to probe all criteria, clinicians engaging in an unstructured interview might probe only for disorders relevant to the initial clinical presentation of the client (Jewel et al., 2004). Such failure to adequately probe may lead clinicians to miss important criteria that lead to a different or additional diagnosis. Furthermore, clinicians may stop probing once they have reached a diagnosis, either because they perceive the interview as complete (Rogers, 2001) or to avoid assigning multiple diagnoses (Weiner et al., 1987).

Following a similar line of reasoning, other researchers have indicated interviewer bias may be a particularly influential factor in the underdiagnosis by unstructured interviews. Based

on their traditional training, Rogers (1986) asserted that most clinicians follow a hypothesis-testing diagnostic model. That is, clinicians test out their clinical hypotheses based on a combination of the clinical presentation, client history, and their own intuition. Such an unstructured approach is highly susceptible to bias. As described by Rogers and Shuman (2000), clinicians "tend to (a) seek and overvalue data that is consistent with their hypothesis, and (b) disregard or under-value data that are inconsistent with their hypothesis" (p. 281). Clinician confirmation-bias can easily lead to either misdiagnosis or missed diagnoses.

In response to these problems, a shift towards more structured and semi-structured interviews has occurred (Groth-Marnat, 2009; Stein, 1987). Structured and semi-structured interviews attempt to ameliorate some of the methodological problems associated with unstructured clinical interviews by systemizing the assessment process through (a) standardized clinical inquires, (b) standardized sequencing, and (c) standardized ratings (Rogers, 2001, 2003). Such standardization ensures that clinical inquires use the same clear and non-pejorative language in a naturally flowing progression (Rogers, 2001). This systematic approach to the assessment of psychological symptoms reduces misdiagnosis (Rogers, 2001). Moreover, each evaluation is adequately comprehensive (Rogers & Wupperman, 2006); thereby reducing missed diagnoses (Rogers, 2001).

Above all methodologically, the standardized nature of structured interviews allows for important systematic comparisons to be made across evaluators, diagnostic groups, and informants (Rogers 2001, 2003). Such standardized comparisons allow for the calculation of diagnostic reliability and validity, introducing a strong component of empiricism to the assessment measure being investigated. Regarding reliability, clinicians can establish the agreement for diagnoses across evaluators (i.e., interrater reliability) or specified time intervals

(i.e., test-retest reliability; Rogers, 2001). In terms of diagnostic validity, Rogers (2003) argued structured interviews are key in establishing discriminant validity—the ability of key symptoms or scales to differentiate between different diagnostic groups—and thereby more precise clinical interpretations.

One salient consideration in assessing adolescents involves the use of collateral sources, particularly parents. While collateral sources often provide additional invaluable information that the adolescents may be unable to provide for themselves, discrepancies between the sources of information are not uncommon (see De Los Reyes & Kazdin, 2005). To illustrate, Cantwell, Lewinsohn, Rohde, and Seeley (1997) assessed the degree of correspondence between 281 parent-adolescent pairs regarding the prevalence of certain adolescent psychopathology. The authors documented substantial variability in their level of agreement, from low (k = .19 for alcohol abuse/dependence) to excellent (k = .79 for conduct disorder). Furthermore, the source reporting the most symptoms varied depending on the disorder being discussed. Adolescents tended to report more internalizing symptoms (e.g., Major Depressive Disorder) while parents generally described a larger number of externalizing symptoms (e.g., ADHD; Cantwell et al., 1997). Logically, this finding makes sense, given that parents can only report on symptoms that can be observed outwardly (i.e., externalizing symptoms) or disclosed by the adolescents.

Jensen and colleagues (1999) further investigated parent-child discrepancies using a large community sample (N = 1,281 parent-child pairs). While the authors corroborated Cantwell et al.'s finding of marked variability in the discrepancies, Jensen et al. found considerably lower concordance rates than the previous study ($M\ k$ of 0.19 versus 0.42, respectively). One possible reason for this difference is that Cantwell et al. (1997) focused on older adolescents (e.g., high-school aged) while Jensen et al.'s (1999) sample covered a greater span in age from 9 to 17. In

partial support of this hypothesis, Jensen et al. (1999) found that older adolescents exhibited slightly higher kappa values than their younger counterparts.

Jensen and colleagues (1999) took their study one step further by investigating truly discrepant cases. That is, they more closely examined the cases in which one informant identified a disorder that the other did not. Predictably, a general pattern emerged in which children were more accurate in identifying discrepant cases of internalizing disorders (e.g., major depression and dysthymia) whereas parents were more accurate with cases of externalizing disorders (e.g., ODD and ADHD). Coupled with Cantwell et al.'s (1997) finding, it is clear that parents and children disproportionately identify various types of psychopathology.

As reviewed by De Los Reyes and Kazdin (2005), such discrepancies have important implications. For example, relying on different informants can lead to differences in conclusions regarding psychiatric diagnosis and/or treatment (see De Los Reyes & Kazdin, 2005). As such, evaluating where discrepancies occur along with the direction and degree of disagreement can be useful (Rogers, 2001). Structured interviews provide the distinctive opportunity to systematically examine discrepancies between sources using a structured and standardized method. As a result, they are frequently used for child and adolescent interviews (Rogers, 2003).

Despite their advantages, structured and semi-structured interviews have several important limitations. Namely, structured interviews require more training and administration time than other assessment methods (Rogers, 2001). Consequently, psychologists continue to seek out ways to obtain such a comprehensive clinical picture in a more time-efficient manner. Multiscale inventories represent one approach to fulfilling these requirements. The subsequent section discusses in detail the evolution of multiscale inventories as an adolescent assessment tool in the context of the most commonly studied and used family of multiscale inventories

(Groth-Marnat, 2009)—the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1943).

Development of the MMPI

Since its original development in 1943, the MMPI has arguably become one of the most easily recognized names in the psychological assessment literature. According to one of the instrument's authors, it was originally designed as an "objective aid in the routine psychiatric case workup of adult patients and as a method of determining the severity of the condition" (Hathaway, 1965, p. 463). Stated differently, the MMPI had the goal of providing health professionals with a broad clinical picture in an objective and easily administered way (Archer, 2005). The authors of the MMPI were not unique in this aim; however, their approach proved to be highly valuable.

Hathaway and McKinley employed empirical criterion keying as a highly innovative approach to the development of MMPI items and scales. As an overview, "empirical criterion keying" refers to selecting and assigning test items to their respective scales based on how well they differentiate clinical groups from normative groups (Archer, 2005; Groth-Marnat, 2009). That is, only items that differentiated a clinical population from a comparison group (e.g., non-clinical community sample) would be selected, regardless of whether the item is theoretically related to the scale. As an illustration, if 20% of a sample of depressed individuals and only 5% of the normative sample responded true to the statement, "I have stomach pains," it could be included in the depression clinical scale. Yet, if similar percentages of the depressed and normative samples answered true, the item would not help to discriminate between groups and

thus would not be included. All of the original clinical scales of the MMPI were established following this procedure (Groth-Marnat, 2009).

In the years following its original development, the MMPI quickly became a widely used assessment measure (Groth-Marnat, 2009). Despite its development for use in a specific context (i.e., adult psychiatric patients), the MMPI has since been applied to a wide range of contexts and populations—including adolescents.

MMPI with Adolescents

In the first formal examination of the MMPI with an adolescent population, Capwell (1945a, 1945b) analyzed the utility of the MMPI scales in discriminating between delinquent and non-delinquent girls. Notably, delinquent girls scored significantly higher than their non-delinquent peers on seven out of the eight MMPI scales, with the only exception being the Hysteria scale. Predictably, based on the delinquent sample, the most pronounced differences were observed on the Psychopathic Deviance (Pd) scale. Capwell (1945a, 1945b) additionally found that these mean differences were stable and persisted in follow-up MMPI administrations to the same sample several months later.

Hathaway and Monachesi (1963) followed in Capwell's (1945a, 1945b) footsteps by further investigating the relationship between the MMPI and delinquency. However, Hathaway and Monachesi improved upon the Capwell's study in two important ways: (a) sample size and (b) depth of information. First, in comparison to Capwell's sample of 186 adolescents, Hathaway and Monachesi cast a much larger net with a sample exceeding 15,000 adolescents from high schools across Minnesota. Second, the authors collected a large wealth of information on each adolescent, including demographic variables (e.g., socioeconomic status), academic information

(e.g., grades), and a variety of personality and behavioral measures (e.g., emotional difficulties). Additional outcome data was further collected on a comparatively smaller but still very large subsample (i.e., 3,976) of the original sample amassed approximately four years later. In addition to a readministration of the MMPI, outcome data on adolescents' conduct behaviors or psychiatric problems were also collected.

Hathaway and Monachesi (1963) determined that several of the MMPI's clinical scales were meaningfully related to juvenile delinquency. In particular, the authors found that elevations on the Pd, Schizophrenia (Sc), and Mania (Ma) scales were predictive of delinquent and antisocial behaviors. Males exhibiting their highest elevations on these scales (either on one scale or a combination of several) had somewhat higher delinquency rates when compared to the entire sample. On the other hand, elevations on the Depression (D), Masculinity-Femininity (Mf), and Social Introversion-Extroversion (Si) scale inhibited these same behaviors. Adolescents with the highest elevations on these scales exhibited a fairly lower delinquency rate (27.1%) as compared to the overall sample (34.6%). Thus, Hathaway and Monachesi (1963) established the utility of the MMPI in predicting juvenile delinquency.

Hathaway and Monachesi's (1963) study also served a larger purpose in allowing the systematic comparison of MMPI patterns and profiles between adolescents and adults. The authors compared the responses of their adolescent sample to those of the original adult inpatient norms. While many differences did appear on the item level, the most striking differences emerged in terms of the specific profile code types. In general, findings indicated that adult code types tended to cluster around scales Hs, D, and Hy, suggesting neurotic presentations (e.g., increased depression and anxiety). In sharp contrast, adolescents produced code types suggesting more psychopathic or psychotic presentations (i.e., elevated scales Pd, Sc, and Ma; Hathaway &

Monachesi, 1963). As an illustrative example, approximately one-quarter of adolescents produced high-points on scale Pd compared to only 10% of adults. Meaningful differences also emerged when comparing non-clinical adults and adolescents, or individuals producing profiles in which none of the clinical scales were elevated. More specifically, nearly one quarter of adult profiles were categorized as within normal limits, whereas less than 10% of their adolescent counterparts fell in this category. This difference becomes even more remarkable when considering this percentage drops below 5% when considering only adolescent males.

For decades following the development of the MMPI, it was common practice to score adolescent MMPI profiles using the original adult norms. Yet, Hathaway and Monachesi's (1963) finding that adolescent response patterns differed significantly from adults on the MMPI underscored the problems associated with this practice. Interestingly, despite this finding, the authors stood firm in their belief that the use of adult norms with adolescents should be continued, stating, "We do not advocate the use of special juvenile norms with the MMPI since to do so would arbitrarily erase much of the contrast between adolescents and adults" (Hathaway & Monachesi, 1963, p. 39).

Still, additional studies highlighted the problems with using adult norms for adolescents. First, adolescent elevations on multiple MMPI scales were inversely related to age (Gynther & Shimkunas, 1966). For example, Gynther and Shimkunas (1966) examined MMPI scales of over 400 psychiatric inpatients ranging in age from 14 to 76. A general trend emerged in which younger respondents exhibited elevated F, Pd, Pa, Sc, and Ma scores. The authors further documented that age accounted for differences in scores above and beyond other proposed demographic variables, such as intelligence, on these same scales (e.g., Pd, Sc, and Ma). In other

words, adolescents appear to produce certain scale elevations as a function of their age and not necessarily actual psychopathology.

As another relevant issue, adolescents consistently endorsed a greater number of items on the MMPI than adults (Archer, 1984; Fontaine, Archer, Elkins, & Johansen, 2001; Marks, Seeman, & Haller, 1974). Specifically, adolescents regularly produced higher elevations on F, Pd, Sc, and Ma scales than either college students (Pancoast & Archer, 1992) or adults (Archer, 1984). Such a stable pattern highlights the pathologizing bias produced by the use of adult norms with adolescents. Archer (2005) made a striking statement in reference to this point:

> It is accurate to state that all samples of adolescents, both from clinical and normal settings, would be expected to show significant elevations on scales Pd, Sc, and Ma, regardless of the actual symptom status of these adolescents, simply and solely as a function of the use of adult norms with adolescents. (p. 259).

The pathologizing tendency towards adolescents has important diagnostic implications in that the use of the MMPI with adolescents frequently led to overdiagnosis of psychopathology. As a cogent example, almost one-third (32.5%) of adolescent psychiatric inpatient admissions were incorrectly diagnosed with psychotic symptoms based on their MMPI profiles scored with adult norms (Chase, Chaffin, & Morrison, 1975).

As an aside, Fontaine and colleagues (2001) outlined several other possible reasons for adolescents' unusual MMPI endorsement patterns. One explanation may be related to the reading level requirement of the instrument, which typically hover around sixth to seventh grade (Johnson & Bond, 1950; Ward & Ward, 1980). While the MMPI was originally intended for use with individuals 16 and older, Dahlstrom, Welsh, and Dahlstrom (1972) suggested the MMPI could be used successfully with children as young as 12, with the caveat that they be intelligent. Thus, it is possible that certain adolescents do not possess the reading proficiency necessary for accurate responding on the MMPI. Moreover, Fontaine et al. (2001) argued that developmental

differences between adolescents and adults may be an important contributing factor. Adolescence is traditionally seen as a turbulent time for youth given the vast amount of change associated with the developmental stage. Adolescents' endorsement of greater impulsivity, rebelliousness, and social isolation may then be consistent with normal development (Archer 1984; Fontaine et al., 2001). Lastly, the authors indicated that the response patterns may be indicative of the tendency of adolescents to be more candid than adults. That is, where adults may be more defensive or engage in greater socially desirable responding, adolescents may be more open and frank in their responses.

Consideration of the problems surrounding the use of adult norms with adolescents led Marks and his colleagues (Marks & Briggs, 1972; Marks, Seeman, & Haller, 1974) to develop the first adolescent-specific norms for the MMPI. Marks and Briggs combined 800 MMPI profiles from the original Hathaway and Monachesi (1963) sample with 1,000 MMPI profiles from adolescents in urban and rural settings across six states in the United States (i.e., Alabama, California, Kansas, Missouri, North Carolina, and Ohio). Norms were reported separately by gender and across four age groups: 14 and younger, 15, 16, and 17. Importantly, Marks et al. (1974) reached another milestone by publishing the first actuarial personality descriptors for MMPI high-point code types in adolescents. The authors developed a code-type classification system by evaluating the personality correlates of over 1,200 adolescents receiving psychotherapeutic services. Overall, correlate information for 29 different codetypes was produced, which allowed for the generation of 61 separate narratives. Taken together, Marks and colleagues' (1974) work represented the first systematic investigation into interpreting adolescent MMPI profiles based exclusively on empirical data obtained from adolescents.

Following the development of adolescent norms, researchers set out to systematically examine the differences in the use of adolescent versus adult norms with adolescent samples (Archer, 1984). As part of Marks et al.'s (1974) research on adolescent norms, the authors compared 952 male and 854 female adolescent profiles against adult norms. They reported that the use of adult norms produced a slight elevation (i.e., within 5*T*) in normal adolescent profiles.

Following this line of research, Klinge and Strauss (1976) found more pronounced differences in a sample of adolescents with more severe psychopathology. They examined the MMPI profiles of 173 adolescents admitted to an adolescent psychiatric service against both adolescent norms and standard adult norms. Similar to Marks et al.'s (1974) findings, Klinge and Strauss noted that the use of adult norms resulted in significantly higher *T* scores than its adolescent counterpart. However, Klinge and Strauss (1976) reported much more dramatic elevations than those reported previously. For example, the authors indicated an elevation of one standard deviation (i.e., 10*T*) on scales Pd and Pt for both genders. Further analyses alarmingly revealed that almost one-quarter of adolescent profiles considered non-psychotic under adolescent norms were classified as psychotic when using adult norms. These conclusions were later replicated in samples of adolescent psychiatric inpatients with and without drug abuse histories (Klinge, Lachar, Grisell, & Berman, 1978) as well as adolescents undergoing intensive treatment (Archer, White, & Orvin, 1979). When considered together, the studies highlight how the choice of norms used to score adolescent MMPIs can dramatically alter its interpretation, especially for adolescents with significant psychopathology.

Additional normative data sets were later developed, in part, due to widespread dissatisfaction with the Marks et al. (1974) norms. Gottesman, Hanson, Kroeker, and Briggs (1987) reanalyzed a portion of Hathaway and Monachesi's (1963) original dataset. The portion

consisted of approximately 14,000 15-year-olds and 3,500 18-year-olds collected from across Minnesota. Their reanalysis of this previously published data included the use of a statistical procedure to normalize the data and thereby minimize the effects of skew in the raw score scale distributions. It also served to increase compatibility in the meanings of *T*-score deviations across scales, such that numerically compatible T-scores on different scales had the same percentile rank (Klinefelter, Pancoast, Archer, & Pruitt, 1990).

Two years later, Colligan and Offord (1989) developed the third set of adolescent norms. The authors administered MMPI protocols to 691 females and 624 males ranging in age from 13 to 17 and from three different states (Minnesota, Iowa, and Wisconsin). Importantly, adolescents reporting a history of psychological treatment or chronic mental disorders were excluded from the normative data set. After concluding that no age effects existed, the authors opted to combine age cohorts while still separating the norms by gender. Of note, Colligan and Offord (1989) followed in Gottesman et al.'s (1987) footsteps by following a normalizing procedure. The authors statistically offset the skew of the raw score distributions and created normalized *T*-scores. This approach is in contrast to the linear *T*-scores used by the Marks et al. (1974) norms (Archer, Pancoast, & Klinefelter, 1989).

Archer et al. (1989) sought to empirically compare the three sets of adolescent norms. Their study had two main objectives: (a) determine the degree to which the three norms converge in terms of MMPI code types and (b) evaluate the effects of using linear versus normalized *T*-scores. The authors examined the MMPI profiles for 300 adolescents between the ages of 12 and 18. The sample combined equal numbers of participants (50 males and 50 females) from each of the following settings: high school, outpatient psychiatric, and inpatient psychiatric. Four MMPI profiles were produced for each participant; specifically, one profile was created from each set of

norms (i.e., Marks et al., Gottesman et al., and Colligan & Offord) with a fourth profile utilizing linear *T*-scores derived from the Colligan and Offord norms. In this way, the second objective of evaluating linear versus normalized T-scores could be examined.

Archer et al. (1989) documented significant variability in terms of scale elevations and code types across the norms. As an example, only approximately half of the 10 most frequent code types in the Marks et al. norms were found among the most frequent code types in either the Gottesman et al. or Colligan and Offord norms. Moreover, the norms also differed in the frequency of within normal limits (WNL) profiles or profiles with no elevations. Interestingly, in several instances the Colligan and Offord norms produced WNL profiles despite significant elevations on scales Pd, Sc, or Ma under the Marks et al. norms. Archer et al. (1989) asserted that such discrepancies are likely partially attributable to differences in the types of *T*-scores utilized (linear versus normalized), but concluded that the era in which the data sets were collected was far more influential.

Other scholars have come to similar conclusions regarding the relative time differences between data collection. The Marks et al. (1974) and Gottesman et al. (1987) norms are based on MMPI protocols administered in the 1960s and 1970s, whereas the protocols from Colligan and Offord (1989) were administered almost two decades later. Scholars (e.g., Williams, 1986) argued that the relatively older age of the Marks et al. (1974) norms hinders accurate interpretation with more contemporary adolescents. As an empirical test of this assertion, Pancoast and Archer (1988) systematically compared the mean MMPI profiles of adolescents collected before 1964 to those of adolescent profiles collected after 1975, using the Marks et al. (1974) norms. Findings supported the use of the Marks et al. norms with profiles collected between 1947 and 1965; these results were expected due to their contemporaneous collection.

Yet, profiles collected after 1975 exhibited significantly different patterns when using the Marks et al. (1974) norms. Specifically, significantly lower mean values were observed on scales L and K along with marked increases in mean scores for F and almost all clinical scales (with the exception of Si). Such conclusions lend support for the use of the more contemporary norms developed by Colligan and Offord (1989).

While the publication of multiple sets of adolescent norms helped to address previous concerns, a number of problems still persisted (Archer et al., 2016). Given the MMPI's original development for use with adults, no items were related specifically to common adolescent experiences (e.g., school difficulties). Moreover, the MMPI did not contain any scales designed to assess for adolescent-specific psychopathology. Additionally, many items contained language that was either considered outdated or inappropriate for use with an adolescent population (e.g., "Sometimes when I was young I stole things." or "My sex life is satisfactory."). Scholars began development of the Minnesota Multiphasic Personality Inventory - Adolescent (MMPI-A; Butcher et al., 1992) with these concerns in mind.

MMPI-A

The Minnesota Multiphasic Personality Inventory - Adolescent (MMPI-A) was published in 1992 with the general goal of standardizing and improving MMPI assessment with an adolescent population (Archer et al., 2016). More specifically, it sought to: (a) compile a national normative sample representative of the general population, (b) create a shorter version of the MMPI, (c) maintain continuity between the original MMPI and the MMPI-A in terms of the preservation of the standard scales, (d) include items directly relevant to adolescent development

and expression of psychopathology, and (e) standardize MMPI assessment with adolescents (Archer, 1992). The MMPI-A accomplished these objectives to varying degrees of success.

With regards to the first objective, normative data on 805 males and 815 females was collected. The sample was relatively diverse in terms of age (ranging from 14 to 18) and geographic region (i.e., California, Minnesota, New York, North Carolina, Ohio, Pennsylvania, Virginia, and Washington). However, it is difficult to conclude that the normative sample is representative of the general adolescent population. While Archer (2005) reported that the ethnic distribution of the sample generally matched U.S. Census figures, Butcher et al. (1992) stated that Hispanic Americans were likely underrepresented. Conversely, adolescents with highly educated parents (e.g., educational level of a bachelors degree or higher) or adolescents living at home with both parents were overrepresented.

MMPI-A authors were more successful at achieving the remainder of the objectives. The final version of the MMPI-A consisted of 478 items (Butcher et al., 1992), a significant reduction from the MMPI 566 items. Despite this reduction, some scholars have argued that the problem with the length of the instrument remains an important concern (Archer et al., 2016). On this point, Archer (2005) indicated that 478 items, which take approximately one hour to complete, could exceed the attention and concentration span of some adolescents. This assertion has important implications for its professional use because decreased concentration and attention could distort or even invalidate MMPI-A results.

Additionally, the authors generally maintained continuity between the original MMPI and MMPI-A. For example, the final version of the MMPI-A contains the original 13 standard scales as well as several of the existing validity scales and supplementary scales (Butcher et al., 1992). Such continuity with the original MMPI was intended to ease the transition for clinicians and

capitalize on the successful aspects of the original measure. Yet, the limitations associated with the original MMPI were also applicable. That is, problems with multidimensionality, content heterogeneity, and extensive item overlap between the scales were also noted on the MMPI-A (Archer et al., 2016). As a result, the MMPI-A exhibited limited discriminant validity (Archer et al., 2016).

Finally, the goal of creating an adolescent-friendly version of the MMPI was mostly achieved. Fifteen content scales were included in the MMPI-A, with four scales being developed uniquely for the MMPI-A (Archer, 1997). These content scales include School Problems, Low Aspirations, Alienation, and Conduct Problems (Butcher et al., 1992). Additional content was included as part of three new supplementary scales: Alcohol/Drug Problem Acknowledgement (ACK), Alcohol/Drug Problem Proneness (PRO), and Immaturity (IMM). Beyond additional items, existing content was revised as necessary for an adolescent population (Butcher et al., 1992).

Additional problems with the MMPI-A surfaced as the new instrument's popularity increased. Of particular concern, Archer (1997) identified the MMPI-A's seemingly low sensitivity, or its ability to identify psychopathology, as the instrument's most important limitation. To illustrate this problem, MMPI-A profiles were frequently classified as WNL despite adolescents being in residential or inpatient psychiatric facilities (e.g., Archer & Elkins, 1999; Baer, Kroll, Rinaldo, & Ballenger, 1999; Fontaine et al., 2001; Forbey, Ben-Porath, & Davis, 2000; Hilts & Moore, 2003). As a review, Hand (2005) amalgamated the existing literature of 32 studies and 5,561 adolescents to determine the mean MMPI-A profile of a clinical sample. He found that the average MMPI-A profile contained no significantly elevated scales. In fact, the highest average *T*-score ($T = 61$) occurred on scale Pd, with all other clinical

scales not exceeding 60*T*. Clearly, a lack of clinical elevations despite exhibiting significant psychopathology has important diagnostic and treatment implications.

As one possible solution, Alperin, Archer, and Coates (1996) hypothesized that a K-correction might increase test sensitivity. To test this hypothesis, the authors compared the accuracy of the non-K-corrected *T*-scores to those using the K correction in their ability to discriminate between two groups: non-clinical and clinical adolescents. The non-clinical group consisted of the 1,620 adolescents in the MMPI-A normative group (described previously) while the clinical group included only 122 adolescents receiving inpatient psychiatric care. Differences between *T*-scores were compared in terms of hit rate, sensitivity, and specificity. Overall, Alperin, Archer, and Coates (1996) concluded that the adoption of a K-correction did not significantly improve test accuracy. Using a threshold of 65*T*, differences between *T* scores were strikingly similar at classifying the clinical sample (K-weighted = 72% versus non-K-weighted = 71%) and the normative sample (K-weighted = 68% versus non-K-weighted = 70%). Thus, the introduction of a K-correction did not prove successful in ameliorating the MMPI-A's low sensitivity.

As another proposal, Fontaine et al. (2001) posited that lowering the threshold for a clinical range elevation might influence sensitivity. Fontaine et al.'s (2001) proposed solution was based on a previous study conducted by Ehrenworth and Archer (1985). Those researchers found that lowering the threshold for clinical range elevations substantially reduced the frequency of WNL profiles by adolescents on the MMPI. Fontaine et al. (2001) applied these findings to the MMPI-A normative sample, comparing the recommended cut score of 65*T* to a reduced cut off of 60*T*. Similar to Ehrensworth and Archer (1985), Fontaine et al. (2001) documented that lowering the clinical threshold did reduce the frequency of WNL profiles.

However, such a reduction came at the cost of reduced overall classification accuracy. Due to the increased possibility of false positives associated with a lowered threshold, Fontaine et al. (2001) recommended that a threshold of 65*T* remain in use.

Archer, Handel, and Lynch (2001) helped to put this difficulty in discriminating between normative and clinical samples into context. According to Archer et al. (2001), a significant number of MMPI-A items do not effectively discriminate between normative and clinical samples. This conclusion was based on a comparison of the item endorsement frequency on the MMPI-A basic scales between the MMPI-A normative sample and two different clinical samples. In general, the basic scales demonstrated poor effectiveness in differentiating the normative from the clinical samples. Most concerning, less than 20% of the items on four scales on the MMPI-A were deemed effective. As one notable exception to the general trend, the Pd scale evidenced effective discrimination between the normative and clinical samples for the majority of its items.

As a point of reference, differences in endorsements were further compared to the item endorsement frequencies of the MMPI-2. Archer, Handel, and Lynch (2001) found that MMPI-A basic scales contained a considerably lower percentage of effective items than those of the MMPI-2. That is, over 80% of MMPI-2 items were deemed effective, while only approximately half of the MMPI-A items (e.g., 56% for boys and 49% for girls) met the same standard. Such findings help to explain, in part, the low sensitivity of the MMPI-A as compared to its predecessor.

The combination of these persisting problems along with other important considerations led the researchers to eventually revise the MMPI-A. Following the wide success of the Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF; Tellegen &

Ben-Porath, 2008, 2011), Archer et al. (2016) applied a similarly innovative approach in developing the MMPI-A's successor: the MMPI-A-RF.

MMPI-A-RF

The MMPI-A-RF (Archer et al., 2016) is the most recently published assessment test in the long line of MMPI measures. The MMPI-A-RF represents a marriage between the content of its predecessor, the MMPI-A, and the innovative development of the MMPI-2-RF. In this way, the MMPI-A-RF takes advantage of the success of the MMPI-A, while still addressing its previously illuminated on problems, namely the problem of heterogeneity (Archer et al., 2016).

The effective integration of the MMPI-A with the MMPI-2-RF occurred in two steps. First, it was necessary to identify the corresponding items between the MMPI-2-RF and MMPI-A (Archer et al., 2016). This step was accomplished by mapping all of the MMPI-2-RF reconstructed scale (RC) items onto the MMPI-A items. The corresponding items formed the initial seed scales for the MMPI-A-RF restructured scales. Next, the scales were further refined through statistical item analyses (i.e., corrected item-total correlations) in addition to a review of item content (Archer et al., 2016). In the final steps, all RC seed scales were correlated with each MMPI-A item. Such correlational analyses allowed for the addition of 19 MMPI-A specific items to be included on MMPI-A-RF reconstructed clinical (RC) scales. Thus, the MMPI-A-RF follows the development and structure of the MMPI-2-RF while still including adolescent-specific content.

Beyond development, the MMPI-A-RF also mirrors the MMPI-2-RF regarding its structure. Akin to the MMPI-2-RF, the MMPI-A-RF contains a separate measure of demoralization (Archer et al., 2016). The demoralization factor was identified as the common

nonspecific factor of distress that is shared among all of the MMPI-A clinical scales. Given its shared variance among the clinical scales, it was predictably found to be a significant contributor to high inter-correlations among the MMPI-A scales (Archer et al., 2016). Furthermore, the MMPI-A-RF employs the three-tiered approach to interpretation associated with the MMPI-2-RF. Starting at the top, the Higher-Order (H-O) Scales provide the broadest level of assessment, followed by the Restructured Clinical (RC) Scales, and the Specific Problems (SP) Scales at the bottom representing the more nuanced and fine level of assessment.

The content of the MMPI-A-RF consists entirely of MMPI-A items. As a research advantage, all previously collected data on the MMPI-A can be conveniently rescored as MMPI-A-RFs (Archer et al., 2016). In fact, the MMPI-A-RF normative sample directly stems from its predecessor's normative sample. A subset of the original sample of junior high and high school students' MMPI-A protocols was rescored as MMPI-A-RFs to form the norms. Moreover, previous research on the MMPI-A may contribute to the burgeoning MMPI-A-RF literature. Thus, the MMPI-A-RF manual presents the psychometric properties and clinical correlates of the MMPI-A-RF from several MMPI-A data sets, representing multiple settings and research methods. In the following section, different research methods will be explored within the context of the currently published clinical correlates of the MMPI-A-RF.

Methodological Considerations in Validating Clinical Correlates

Clinical correlates are an essential part of the external validation of any newly developed multiscale inventory. Such correlates provide additional meaning to the test scores by establishing the specific relationship between the measure itself (e.g., MMPI-A-RF scale scores)

and some behavior or characteristic of the examinee. In this way, the clinical correlates provide some context for the scale elevations.

Research Methodology

The MMPI-A-RF manual provides detailed information on the external correlates of each MMPI-A-RF scale. Such correlates stem from three external MMPI-A data sets: Forbey and Ben-Porath (2003), Veltri et al. (2009), and Handel et al. (2011). Importantly, these data sets differ in settings as well as sample demographic characteristics (see Table 1). For example, they differed in terms of ethnicity: Veltri et al.'s (2009) being mostly European American and Handel et al.'s (2011) being mostly African American. As discussed next, the most important difference involved the methodology utilized.

Regarding methodology, Handel et al. (2011) used a battery of symptom and behavior checklists for clinical correlates. The measures were comprised of the Child Behavior Checklist (CBCL; Achenbach & Edelbrock, 1983), Youth Self-Report (YSR; Achenbach & Rescorla, 2001), and Disruptive Behavior Rating Scale (DBRS; Barkley & Murphy, 1998). Symptom and behavioral checklists provide ratings of an adolescent's functioning as described by either a caregiver (i.e., CBCL and DBRS) or the evaluated adolescent (i.e., YSR). As an important criticism, symptom and behavioral checklists tend to be limited in scope for a variety of reasons. For example, some checklists, such as the DBRS, assess only one or two aspects of psychopathology. On the other hand, those checklists that attempt to assess a broader range of psychopathology (i.e., CBCL) are limited by their brevity of measured constructs. By prioritizing brevity over breadth, only a limited number of clinical correlates can be assessed.

Table 1

Characteristics of the MMPI-A-RF Clinical Correlate Studies

Sample		*n*	Age (*M*)	Ethnicity (%)			Correlate Measures
				European American	African American	Other	
Handel et al. (2011)							
Forensic	Males	326	15.81	26.4	63.2	10.4	Symptom checklists
	Females	195	15.91	26.7	64.1	9.1	
Veltri et al. (2009)							
Psychiatric Inpatient	Males	86	16.10	94.2	3.5	2.4	Record review
	Females	216	15.82	88.0	7.4	4.7	
Forensic	Males	199	16.01	-	-	-	Record review
Forbey and Ben-Porath (2003)							
Residential Treatment	Males	258	14.83	74.0	19.8	6.3	Clinician ratings
	Females	114	14.86	68.4	24.6	7.0	

Note. This table has been adapted from Archer et al. (2016). The "Other" ethnicity group refers to any ethnicity other than European or African American, such as Hispanic or Biracial Americans.

As a different methodology, Veltri et al. (2009) relied on information coded from adolescents' records. External variables were thus derived from sources such as intake interviews, parent interviews, psychiatrist reports, and medical/legal records. Veltri et al. (2009) argued that the record review variables served as a strength for their study because they stemmed from multiple sources and not solely self-report. As a strength, the authors developed a specialized record review to standardize the collection of relevant variables. While an important step, inter-rater reliability across variables ranged markedly. As an example, kappa coefficients ranged from a very low of .23 ("history of lying" in the forensic sample) to a perfect 1.00 ("hopelessness" in the inpatient sample). Obviously low inter-rater reliability estimates limits its correlates' clinical utility. Moreover, much like with checklists, the sole reliance on variables coded from records comes at the cost of breadth of coverage with many characteristics of adolescent psychopathology and personality being omitted. Veltri et al. (2009) admitted that additional clinical variables would have been useful, such as disordered eating, somatic complaints, and racing thoughts.

Forbey and Ben-Porath (2003) utilized a third approach to external correlates by developing their own clinician-rated measure. Their measure, the Adolescent Client Description Form (ACDF), is a 161-item instrument composed of personality and symptomatic descriptors rated by the adolescent's primary clinician on a 5-point Likert-type scale (Forbey & Ben-Porath, 2003). The ACDF proved to be a better method than record review for several reasons. First, the ACDF represents a more reliable measure than the record review form utilized by Veltri et al. (2009). Specifically, the range of interrater reliabilities across the scales of the ACDF (.69-.91) was generally stronger than Veltri et al.'s (2009) record review form (.23-1.00). Second, the ACDF utilized the adolescent's primary clinician, who is presumably very familiar with the

particular psychopathology and personality characteristics of the adolescent and would be able to accurately report his or her functioning. Third, the ACDF assesses a much wider breadth of clinically relevant information than the other two approaches discussed. However, its chief limitation was its development specifically for this study, and thus, it has not been independently validated (Forbey & Ben-Porath, 2003).

Notably, none of the previous MMPI-A-RF correlate studies employed structured interviews as a systematic means of establishing clinical correlates. As discussed previously, structured and semi-structured interviews address two important limitations associated with other assessment measures. First, they standardize how and what information is gathered, thereby providing a systematic evaluation of reliability and reducing criterion variance. Second, they are more comprehensive, thus missing fewer diagnoses or other clinically relevant information (Rogers, 2001). Relatedly, structured interviews utilize mental health professionals trained to evaluate symptoms and their severity. Thus, the current dissertation contributes to the MMPI-A-RF clinical correlate literature via the application of a semi-structured diagnostic interview paradigm.

The current dissertation selected the following well-validated interview for external correlates: Kiddie Schedule for Affective Disorders and Schizophrenia-Present and Lifetime Version (K-SADS-PL; Kaufman et al., 1997). The K-SADS-PL is a semi-structured diagnostic interview developed for use specifically with children and adolescents. The K-SADS-PL offers two major advantages in the evaluation of clinical correlates. First, the K-SADS-PL provides detailed information on not only the presence of symptoms but also their severity. In this way, the strength of clinical correlates can be systematically assessed. Second, the K-SADS-PL has been found to be a highly reliable instrument in assessing psychopathology in youth (Kaufman et

al., 1997). Taken together, the K-SADS-PL proves to be a useful measure in providing reliable and detailed clinical information for the formulation of clinical correlates.

Given these advantages, previous clinical correlate studies have also utilized the K-SADS. As one relevant example, Cashel, Rogers, Sewell, and Holliman (1998) explored the clinical correlates of the MMPI-A in a sample of male juvenile offenders. While utilizing en earlier version than the K-SADS-PL, Cashel et al. (1998) systematically compared the MMPI-A Basic Scales with symptoms and diagnoses derived from the K-SADS-PL's predecessor, the Schedule of Affective Disorders and Schizophrenia for School-Age Children (K-SADS-III-R; Ambrosini, Metz, Prabucki, & Lee, 1989). As expected, findings demonstrated significant correlations between MMPI-A scales and conceptually relevant items from the K-SADS-III-R. Thus, Cashel et al. (1998) lent general support for the convergent validity of the then newly developed MMPI-A.

Gendered vs Non-Gendered Norms

Another important consideration related to methodology involves whether clinical correlates should be gender-specific. Traditionally, psychological assessment measures have employed gender-specific norms. This separation follows the research literature's finding of gender differences in psychopathology and personality traits (Hathaway & Monachesi, 1963; Mason, Bubany, & Butcher, 2012). However, the publication of the MMPI-2-RF, and later the MMPI-A-RF (Archer et al., 2016), broke with this tradition with non-gendered norms. Predictably, the MMPI-A-RF's use of non-gendered norms again brought the gendered versus non-gendered norms debate back into the spotlight.

Many scholars have expressed criticism of the MMPI-2-RF's implementation of non-gendered norms. Those opposing non-gendered norms (e.g., Butcher & Williams, 2012) claim that this practice may result in possible gender bias. That is, important gender differences will be obscured if gendered norms are not utilized. For example, Butcher (2010) explicitly stated:

> Unlike the original MMPI and MMPI-2, in which separate gender norms were provided, the MMPI-2-RF authors combined the comparison sample. This situation may result in different standards being applied for men and women in assessment and prediction. Further study of this potential bias needs to be conducted. However, the MMPI-2-RF manuals do not provide the information necessary for exploring this question because raw score data by gender are not reported. (p. 14)

Beyond stating his concern, Butcher (2010) called for continued research to examine the effects of the non-gendered approach.

In stark contrast, scholars supporting the use of non-gendered norms argue that gender differences are actually more apparent under non-gendered norms (Archer et al., 2016; Ben-Porath & Fens, 2012). According to Ben-Porath and Flens (2012), the non-gendered norms hold the scores of males and females to the same standard, thereby preserving gender differences. Moreover, the means and standard deviations of the MMPI-2-RF scales are still reported by gender (Ben-Porath & Tellegen, 2008). Thus, gender-specific analyses can still be performed. Looking across sides, it is clear that the debate between gendered versus non-gendered norms continues to be contested.

Coinciding with the debate, many studies have examined the impact of gender on the various versions of the MMPI with adolescents. Dating back to when the original MMPI was routinely used with adolescents, Williams and Butcher (1989) concluded that any gender differences across scales were small and insignificant. Yet, other studies documented different clinical correlate patterns for males and females (Wrobel & Lachar, 1992). As a cogent example, Wrobel and Lachar (1992) recognized different behavioral manifestations for elevations on the

Psychasthenia (Pt): girls were described as acting out and suicidal whereas boys tended to lack self-confidence. Therefore, Wrobel and Lachar (1992) argued for the continued examination of the effect of gender on clinical correlates.

Similar sentiments were also expressed regarding the MMPI-A. Archer (1997) called for the evaluation of the effect of gender on clinical correlates after documenting significant gender differences in terms of scale elevations on the MMPI-A. Many studies answered Archer's call. For example, Handel et al. (2011) found that females' scores on scale D were more predictive of external variables than those for boys. Given the consistent pattern of gender differences on both the MMPI and MMPI-A, it stands to reason that similar differences would also be present on the MMPI-A-RF. More importantly, it is critical to understand how such gender differences may affect interpretation. As a result, the current dissertation investigates the effect of gender on the MMPI-A-RF in terms of psychometric properties, clinical correlates, and interpretation.

Current Study

Given its recent publication date, the MMPI-A-RF empirical literature is only beginning to flourish. The current study served to build on previous-MMPI-A-RF studies by evaluating its clinical correlates in a sample of mental health referred youth. In addition, the current study helped determine the psychometric properties of the MMPI-A-RF in an adolescent offender and partially hospitalized youth sample in terms of reliability, convergent validity, and discriminant validity.

As part of its initial validation, MMPI-A-RF clinical correlates were established (Archer et al., 2016). As previously reviewed, the correlate studies have employed extra-test indicators including adjective or behavior checklists (Handel et al., 2011), adolescent records (Veltri et al.,

2009), or information from clinician ratings (Forbey & Ben-Porath, 2003). This dissertation evaluates the clinical correlates of the MMPI-A-RF using a well-validated semi-structured interview, the K-SADS-PL.

The K-SADS-PL is a semi-structured interview used to assess for psychopathology in children and adolescents. It can provide clinically relevant information on a variety of diagnostic groups. In the current study, the ten most-relevant diagnostic sections were administered: depressive disorders, mania, psychosis, panic, generalized anxiety disorder, attention-deficit/hyperactivity disorder (ADHD), oppositional defiance disorder (ODD), conduct disorder (CD), alcohol use, substance use, and posttraumatic stress disorder (PTSD). These diagnostic sections were chosen specifically due to their relevance in the current sample. Substance use and externalizing disorders (e.g., ODD and CD) sections were included given documented prevalence rates exceeding 85% in certain samples of juvenile offenders (Dixon, Howie, & Starling, 2004, 2005; Tarolla, Wagner, Rabinowitz, & Tubman, 2002). Substantial rates (e.g., exceeding 50%) of mood disorders (Dixon et al., 2004) and exposure to trauma (Cauffman et al., 1998) have also been documented in juvenile offender samples. As a result, the depressive disorders and PTSD sections of the K-SADS-PL were also administered. While relatively far fewer percentages of juveniles are diagnosed with ADHD, psychosis, mania, and Generalized Anxiety Disorder, these sections were still included in the current study in order to ensure the inclusion of a variety of psychopathology.

Further, the current study adds to the current literature by empirically investigating how gender might influence clinical interpretations of psychopathology. As discussed previously, scholars (Archer et al., 2016; Williams & Butcher, 1989) have debated whether gendered or non-gendered norms are most appropriate for multiscale inventories. Ultimately, the answer to this

conflict has important theoretical implications as well as direct relevance to clinical interpretations.

As a supplementary aim, the current dissertation also examines how the MAYSI-2 (Grisso & Barnum, 2006) performs in a mental health referred sample of juveniles. Given the MAYSI-2's frequent use in juvenile justice settings (Grisso, 1999), it is imperative to empirically test its effectiveness in detecting psychological difficulties. According to Grisso (1999), the MAYSI-2 is intended to identify behaviors and feelings related to symptoms of mental illness. Thus, the current study aimed to contribute to the current literature by formally testing the relationship between the subscales of the MAYSI-2 and current DSM-5 symptomatology.

Research Questions and Hypotheses

Research Question 1: Do the MMPI-A-RF RC scales evidence strong homogeneity in a sample of youth with mental health needs?

Hypothesis 1: The MMPI-A-RF scales would evidence strong homogeneity.

Research Question 2: Do the MMPI-A-RF RC scales evidence construct validity in a sample of youth with mental health needs?

Hypothesis 2: As evidence of convergent validity, the MMPI-A-RF scales would exhibit moderate correlations with the criterion measure, namely the K-SADS-PL.

Hypothesis 3: For discriminant validity, the MMPI-A-RF scales would show lower correlations with other measures of dissimilar psychopathology

Research Question 3: Do the MMPI-A-RF RC scales accurately classify relevant DSM-5 diagnoses?

Hypothesis 4: Restructured clinical scales would predict relevant DSM-5 diagnoses as measured by the K-SADS-PL.

Research Question 4: Are MMPI-A-RF RC scales' interpretive statements related to serious mental disorders?

Hypothesis 5: MMPI-A-RF RC scales would correlate moderately with symptomatology of serious mental disorders as measured by the K-SADS-PL.

Research Question 5: Do non-gendered norms produce accurate clinical interpretations?

Hypothesis 6: Male and female youth would differ in terms of elevations on RC scales of the MMPI-A-RF.

Hypothesis 7: Male and female youth would have different clinical correlates on MMPI-A-RF scales.

Research Question 6: Are MMPI-A-RF Higher-Order scales' interpretive statements related to serious mental disorders?

Hypothesis 8: MMPI-A-RF H-O scales would correlate moderately with serious mental disorders symptomatology as measured by the K-SADS-PL.

Supplementary Research Questions

Supplementary Research Question 1: Are MAYSI-2 subscales effective at screening DSM-5 symptomatology?

Supplementary Research Question 2: Does the absence of clinical elevations on MMPI-A-RF RC scales (i.e., WNL profiles) provide any clinical correlates?

Supplementary Research Question 3: Do elevations on the MMPI-A-RF Antisocial Behavior scale and Behavioral/Externalizing scale predict aggressive conduct problems in youth with mental health needs?

CHAPTER 2

METHOD

Design

The current study examined the psychometric properties and clinical correlates of the MMPI-A-RF using a mixed within-and-between-subjects correlational research design. Broadly, correlations between study measures (i.e., MMPI-A-RF, K-SADS-PL, and MAYSI-2, discussed subsequently) were evaluated as the correlational component of the study. More specifically, correlations were primarily used to evaluate construct validity; see the following section Operationalization of Construct Validity for more details.

Regarding the within-subjects component, all three of the study measures were compared for each participant. Specifically, the within-subjects aspect of the design allowed for the comparison of the MMPI-A-RF RC scales, K-SADS-PL diagnoses and symptoms, and MAYSI scales. For gender differences, the current study employed a between-subjects component with MMPI-A-RF profiles of males and females being systematically compared. This component allowed for the assessment of gender differences, separately for (a) magnitude of elevations and (b) clinical interpretations on MMPI-A-RF scales.

Operationalization of Construct Validity

For construct validity, the MMPI-A-RF RC scales were assessed on two levels: (a) item level and (b) scale level. At the item level, MMPI-A-RF items were expected to correlate higher with its assigned scale than the other RC scales. As summarized in the next several paragraphs, clinical correlate research has adopted several standards that vary in their methodological rigor.

Zimmerman and Mattia (2001) utilized stringent criteria for construct validity in their validation of the Psychiatric Diagnostic Screening Questionnaire (PDSQ; Zimmerman 2002,

Zimmerman & Mattia, 1999, 2001). First, the authors specified that items must evidence good convergent validity, which they defined as evidencing correlations greater than .30 with its designated scale. Additionally, Zimmerman and Mattia (2001) required items to also demonstrate discriminant validity, which they quantified as correlations at least .20 higher on its designated scale versus other scales.

Siefert, Sinclair, Kehl-Fie, and Blais (2009) adopted a much more relaxed approach to construct validity in their investigation of the Personality Assessment Inventory (PAI; Morey, 1991, 1997). They similarly argued items needed to evidence a significantly greater correlation with its designated scale, as compared to its non-designated clinical scales, but did not quantify a minimum magnitude for this difference. Moreover, Siefert et al. (2009) described calculating cross-scale correlations to supplement scaling success rates. That is, they advised finding the percentage of items producing a correlation with their respective scale lower than item-scale correlations with at least one other scale. For the sake of completeness, the current study analyzed item-level construct validity using both the Zimmerman and Mattia and Siefert et al. standards.

Consistent with previous clinical correlate research, construct validity was also assessed on a scale level. That is, convergent and discriminant validity was determined by inspecting correlations between MMPI-A-RF RC scales and K-SADS-PL item's symptom severity scores and diagnostic section total scores. Unfortunately, previous clinical correlate research has varied considerably in the benchmarks for classifying effect sizes for such correlations. For example, recent studies have classified correlations as low as .20 as "significant" (Handel et al., 2011; Veltri et al. 2009) or a "medium effect size" (Gignac & Szodorai, 2016). In sharp contrast, Rogers and colleagues utilized a far more rigorous standard: .35 for moderate, .53 for large, and

.60 for very large (Rogers, 2008; Rogers & Bender, 2013; Rogers, Williams, Winningham, & Sharf, in press).[1] For the analyses, the current study opted to use the more stringent criteria set forth by Rogers et al.

Participants

The current sample consisted of juveniles between the ages of 13 and 18.[2] All juveniles were youth with mental health needs, either in a detention or partial hospitalization setting. In the current study, all legally involved participants were youth residing in a juvenile detention center post-adjudication, while partially hospitalized youth were youth enrolled in a program in which they participated in intensive outpatient psychological services during the day, but returned home each night.

Inclusion criteria were kept minimal to increase the representativeness of the sample. Besides fluency in English, appropriate assent and parental consent were required. For exclusion criteria, adolescents were excluded if they had lower than fourth grade reading levels, given the fourth grade reading level of the MMPI-A-RF (Archer et al., 2016). Participants were not excluded based on race, gender, or sexual orientation.

Study Measures

Kiddie Schedule for Affective Disorders and Schizophrenia-Present and Lifetime Version (K-SADS-PL)

The K-SADS-PL (Kaufman et al., 1997) is a semi-structured diagnostic interview

[1] A correlation of .35 is equivalent to a Cohen's *d* of 0.75 (Rogers, 2008; Rogers et al., in press); .53 is equivalent to a Cohen's *d* of 1.25 (Rogers, 2008; Rogers et al., in press); .60 is equivalent to a Cohen's *d* of 1.50 (Rogers, 2008; Rogers et al., in press).

[2] The MMPI-A-RF is intended for use with ages 14 through 18; however, Archer (2017) specifies 13-years-old may be evaluated if they meet certain criteria (e.g., appropriate reading level).

designed to assess for present and lifetime prevalence of psychological symptoms in children and adolescents. Most psychological symptoms are rated on a four-point Likert scale: 0 (no information available), 1 (not present), 2 (sub-threshold) or 3 (meets threshold). Ratings can be assessed on a symptom-level basis as well as aggregated to generate clinical diagnoses.

The K-SADS-PL has demonstrated good psychometric properties in terms of interrater reliability, test-retest reliability, and concurrent validity. For example, Kaufman and colleagues (1997) reported interrater agreement exceeding 90% for both current and lifetime diagnoses. Furthermore, Kaufman et al. (1997) reported test-retest reliability estimates ranging from moderately good (k = .63 for ADHD) to excellent (k = .90 for Major Depressive Disorder). Regarding concurrent validity, Kaufman and colleagues (1997) reported positive results when comparing K-SADS-PL scores to other rating scales of psychopathology (e.g., Child Behavior Checklist, Achenbach & Edelbrock, 1983; Beck Depression Inventory, Beck et al., 1961; Children's Depression Inventory, Kovacs, 1985).

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

The MAYSI-2 (Grisso & Barnum, 2006) is a psychological measure used specifically in juvenile justice samples to screen for a broad range of psychopathology. Its 52 yes-no items are organized into seven scales: Alcohol/Drug Use, Angry-Irritable, Depressed-Anxious, Somatic Complaints, Suicide Ideation, Thought Disturbance, and Traumatic Experiences. The MAYSI-2 scales range in internal consistency from good (.73) to excellent (.86). In terms of validity, Archer, Simonds-Bisbee, Spiegel, Handel, and Elkins (2010) found good convergent validity with relevant clinical variables (e.g., history of suicidal behavior, drug use, or psychiatric placements).

Minnesota Multiphasic Personality Inventory - Adolescent Restructured Form (MMPI-A-RF)

The MMPI-A-RF (Archer et al., 2016) is a multiscale inventory designed specifically for clinical use with youth aged 14 to 18 years. The MMPI-A-RF consists of 241 true-false questions that comprise 48 scales: 6 validity scales, 3 higher-order scales, 9 restructured clinical scales, 25 specific problem scales, and 5 personality psychopathology five (PSY-5) scales. Overall, the MMPI-A-RF provides information in terms of over-arching dimensions (i.e., Emotional Internalizing Dysfunction [EID], Thought Dysfunction [THD], and Behavioral/Externalizing Dysfunction [BXD]) as well as more specific functioning in the somatic/cognitive, internalizing, externalizing, interpersonal, and interest domains.

The MMPI-A-RF has demonstrated generally acceptable reliability in terms of internal consistency and test-retest reliability. Alpha coefficients ranged from .45 (RC9) to .80 (RCd) for males in the standardization sample and .52 (RC9) to .83 (RCd) for females. Test-retest correlations averaged .68 for the restructured clinical scales, ranging from .56 (RC8) to .82 (RCd; Archer et al., 2016). Convergent and discriminant validity was established with a wide range of extra-test indicators such as record review, Adolescent Client Description Form, Child Behavior Checklist, and Youth Self-Report (Archer et al., 2016).

Wide Range Achievement Test – Fourth Edition (WRAT-4) – Word Reading Subtest

The Wide Range Achievement Test – 4th edition (WRAT-4; Wilkinson & Robertson, 2006) is a brief, standardized assessment of academic achievement. The Word Reading subtest provides a basic estimate of individuals' reading level. It has been successfully used in a variety of professional settings (e.g., Davis, Michielutte, Askov, Williams, & Weiss, 1998). The WRAT-4 word reading subtest has demonstrated excellent internal consistency reliability (median α =

.98). In addition, it is strongly correlated with the WIAT-II reading composite scores (r = .78; Wilkinson & Robertson, 2006).

Procedure

Recruitment and Informed Consent

Participants were recruited from Collin County Juvenile Probation or Changes Outpatient of Frisco. As a first step, Collin County Juvenile Probation[3] or Changes Outpatient of Frisco staff provided the researcher with a list of names and contact information of all juveniles in their facilities. The researcher contacted the parents/guardians of all juveniles with study information as part of the consent process. They were then given the opportunity to ask the researcher any questions about the study and to consent or decline their child's participation in the study.

Once parental consent was completed, the researcher asked each individual youth if he or she would like to participate in the study. They were provided study information and any questions were answered. Informed assent was obtained from each participant prior to his or her involvement in the study.

Standardized Measures

First, demographic information was obtained from participants. To determine suitability for the study, the WRAT-4 was administered to assess each individual's reading level. If the juvenile's reading abilities were determined to be below a fourth grade reading level, the participant was thanked and excused. Assessing a potential participant's reading level in the beginning of the study was critical as it ensured that all participants were able to successfully

[3] Initially, a passive consent procedure was implemented at the juvenile detention center. However, approximately one month into the study, a change in administration policy required an active consent process, as detailed here.

comprehend and accurately complete all study measures.

For participants continuing in the study, the K-SADS-PL was administered next to develop rapport with each participant individually. As described previously, ten sections of the K-SADS-PL—particularly relevant to an adolescent clinical population—were administered in the following order: depressive disorders, mania, psychosis, panic, generalized anxiety disorder, attention-deficit/hyperactivity disorder, oppositional defiant disorder, conduct disorder, alcohol use, substance use, and posttraumatic stress disorder.

After administration of the K-SADS-PL, juveniles completed the final two study measures. Each juvenile completed the MMPI-A-RF and the MAYSI-2. Importantly, the two self-report measures were completed in a counter-balanced order to minimize any ordering effects.

Manipulation Check and Debriefing

At the conclusion of testing, juveniles were asked, "How would you rate your level of effort in answering my questions on a scale of 1 to 10, where 1 means you were dishonest and 10 means you were open and honest throughout the study?" This assessment served as the current study's manipulation check and ensured that participants reported at least adequate effort (i.e., reported effort greater than 5).

Participants were also debriefed, which included a review of the general description of the study and its aims. They were given time to ask about the study in order to ensure a full understanding.

CHAPTER 3

RESULTS

Refinement of the Sample

The original sample consisted of 68 juveniles. A total of two participants were subsequently excluded from the final sample. First, one participant was removed from subsequent analyses due to an incomplete MMPI-A-RF, the primary study measure. A second participant was excluded from data analyses due to poor effort; he exceeded established cut scores on multiple validity scales of the MMPI-A-RF for genuine responding. A third exclusion rule involved removing participants rating themselves as having given less than moderate effort (i.e., < 5 out of 10) in answering study questions honestly; however, no participants met this criteria. To the contrary, study participants consistently rated their effort as high (M = 9.53, SD = 0.69, ranging from 7 to 10) during the manipulation check. Finally, participants falling below the study's fourth grade reading level requirement were excluded. No participants were removed under this criteria.

Description of the Final Sample

The final sample was comprised of 66 adolescents. Ages of participants were well represented across the span of adolescence, ranging from age 13 to 18. On average, participants were 15.64 years old (SD = 1.53) with more females (60.6%) than males (39.4%). In addition, participants represented a range of self-identified ethnicity, with the majority of the sample identifying as European American (69.7%). The remainder self-identified as follows: 13.6% as biracial or multiracial, 7.6% African American, 6.1% Hispanic American, 1.5% Native American, and 1.5% Asian American.

With reference to academic abilities, participants averaged at the beginning of the tenth grade (M = 10.08, SD = 1.61). Their reading level, as measured by the WRAT-4, was very similar to their academic achievement (M = 10.44, SD = 2.54). Interestingly, although participants' average grade and reading level were commensurate, a much larger range was observed in their reading levels, which spanned 10 grade levels, from almost fifth grade to college-level reading.

As stated in the Methods, the study recruited participants from two different settings to more fully examine how the MMPI-A-RF performs with specialized populations. A juvenile detention center was used to investigate the MMPI-A-RF in a juvenile justice setting,[4] whereas a partial hospitalization program provided a more intensive treatment setting. While salient clinical symptoms were seen in both of these settings, it is critical to examine how these sites differ in other important ways.

Regarding ethnicity, significant differences emerged related to research setting (see Table 2). While both samples consisted mainly of European Americans, they differed with respect to which other ethnicities were represented. By comparison, the juvenile detention setting was far more ethnically diverse (Cramer's V = .51, p = .01) than the partial hospitalization setting. More specifically, more than half (54.6%) of the detention sample was composed of minority ethnic groups, about double of the partial hospitalization setting (25.5%). Looking more closely, 36.4% of the juvenile detainees identified as African American, whereas only one participant (1.8%) in the partial hospitalization program did so.

[4] As previously discussed, only a small number of juvenile offenders (i.e., 11) were tested following a change in administration policy.

Table 2

Gender, Ethnicity, and Language Differences between Partial Hospitalization and Juvenile Detention Settings

	Total sample		Setting				X^2	*p*
			Partial Hospitalization		Juvenile Detention			
	n	%	*n*	%	*n*	%		
Gender							1.27	.21
Male	26	39.4	20	36.4	6	54.5		
Female	40	60.6	35	63.6	5	45.5		
Ethnicity							16.95	.01
European American	46	69.7	41	74.5	5	45.5		
African American	5	7.6	1	1.8	4	36.4		
Hispanic American	4	6.1	4	7.3	0	0.0		
Biracial	9	13.6	7	12.7	2	18.2		
First Language							5.81	.06
English	63	95.4	54	98.2	9	81.8		
Other	3	4.5	1	1.8	2	18.2		

The adolescents also varied considerably across several other demographic variables as a function of their respective sites (see Table 3). As fully expected, they significantly differed regarding their involvement with the juvenile justice system. Predictably, juvenile detainees tended to have more than five arrests, a far cry from their clinical counterparts, for whom 87.3% reported no arrests. Such marked differences produced a very large effect size (d = -2.54).

Age and grade levels remain virtually identical across research settings. Interestingly, although the groups did not differ in years of education, the detainees predictably faced more achievement challenges. They lagged behind their partially hospitalized counterparts by approximately one grade in reading levels (d = 0.39).

Table 3

Demographic Differences between Partial Hospitalization and Juvenile Detention Settings

	Setting								
	Total sample (N = 66)		Partial Hospitalization (n = 55)		Juvenile Detention (n = 11)		*F*	*p*	*d*
	M	*SD*	*M*	*SD*	*M*	*SD*			
Age	15.64	1.53	15.62	1.60	15.73	1.10	.05	.83	-0.07
Education	10.08	1.61	10.07	1.71	10.09	1.04	.01	.97	-0.01
Reading Level	10.44	2.54	10.60	2.55	9.62	2.43	1.38	.24	0.39
Number of Arrests	1.27	2.99	0.35	1.27	5.81	4.57	61.49	**<.001**	-2.54

Participants reported experiencing significant mental health difficulties, regardless of research setting. Nonetheless, the specific mental health symptomatology demonstrated between research settings varied considerably (see Table 4). The average severity of internalizing symptomatology appears strikingly identical across settings. Yet, upon further examination, it is apparent that looking broadly overlooks some key diagnostic differences. For example, adolescents in the partial hospitalization program exhibited more severe depressive symptoms than juvenile offenders, to a very large degree (d = 1.38). Commonly comorbid anxious symptomatology was also more severe in partially hospitalized adolescents, although to a much more moderate degree ($M\ d$ = 0.43). With regard to PTSD, a trend emerged in which juvenile detainees manifested more severe PTSD symptoms (d = -0.58). The frequency of these observed symptoms likely reflects the high prevalence of trauma and PTSD in detained youth (Dierkhising, Ko, Woods-Jaeger, Briggs, Lee, & Pynoos, 2013).

Not surprisingly, psychosis appeared to be underrepresented across both partial hospitalization and juvenile detention settings. In fact, less than 10% of each sample (i.e., 5

juveniles in partial hospitalization setting and 1 in juvenile detention) exhibited salient psychotic symptoms. Such a low prevalence rate was expected given the relative infrequency of psychosis during adolescence in general (see Kelleher, Connor, Clarke, Devlin, Harley, & Cannon, 2012).

Table 4

Differences in Mean Symptom Severity by Common Adolescent Disorder between Partial Hospitalization and Juvenile Detention Settings

K-SADS-PL		Setting								
		Total sample (N = 66)		Partial Hospitalization (n = 55)		Juvenile Detention (n = 11)		F	p	d
		M	*SD*	*M*	*SD*	*M*	*SD*			
Internalizing	MDD	1.71	0.33	1.78	0.29	1.37	0.33	17.92	**<.001**	1.38
	Mania	1.31	0.60	1.27	0.64	1.53	0.30	1.72	.19	-0.44
	Psychosis	0.25	0.41	0.22	0.38	0.38	0.55	1.48	.23	-0.39
	Panic	1.34	0.65	1.39	0.70	1.13	0.32	1.34	.25	0.40
	GAD	1.62	0.33	1.65	0.32	1.50	0.37	2.03	.16	0.46
	PTSD	1.00	0.71	0.93	0.72	1.33	0.51	3.02	.09	-0.58
	M	1.21	0.30	1.21	0.31	1.21	0.28	0.00	.99	0.00
Externalizing	ODD	1.62	0.66	1.53	0.66	2.08	0.38	7.14	**.01**	-0.88
	CD	1.94	0.73	1.79	0.64	2.66	0.73	16.19	**<.001**	-1.33
	Alcohol	0.60	0.81	0.47	0.73	1.16	0.97	7.12	**.01**	-0.89
	Drug	0.73	1.00	0.52	0.87	1.70	0.97	15.92	**<.001**	-1.33
	ADHD	1.75	0.55	1.73	0.56	1.87	0.50	0.59	.45	-0.25
	M	1.31	0.53	1.18	0.43	1.90	0.58	20.80	**<.001**	-1.58

Note. Scores $\geq$ 2 are considered subthreshold. Scores $\geq$ 3 are considered threshold. MDD = Major Depressive Disorder; GAD = Generalized Anxiety Disorder; PTSD = Posttraumatic Stress Disorder; ODD = Oppositional Defiant Disorder; CD = Conduct Disorder; Alcohol = Alcohol Use Disorder; Drug = Substance Use Disorder; ADHD = Attention Deficit/Hyperactivity Disorder.

As expected, juvenile detainees reported vastly more severe externalizing symptoms than those in the partial hospitalization program (d = -1.58). Moreover, with the exception of ADHD,

all differences between samples resulted in large to very large effect sizes (*M d* = -1.11). Diagnostic considerations easily explain these marked differences. For example, many of the inclusion criteria for a diagnosis of conduct disorder (e.g., fire setting or vandalism) represent acts that would presumably lead youth to become involved with the legal system and eventually be detained.

Together, the two data collection settings represent a diverse range of adolescent psychopathology. Such range of psychopathology is critical in order to adequately assess clinical correlates. Therefore, despite some, mostly expected, differences between the samples, the samples were combined for all subsequent analyses in an effort to increase representativeness of significant mental health symptomatology.

Research Questions and Hypotheses

Research Question 1: Do the MMPI-A-RF RC scales evidence strong homogeneity in a sample of youth with mental health needs?

In their classic work, Clark and Watson (1995) set forth specific guidelines to ensure strong homogeneity and reliability for new scales. Specifically, the authors argued that new scales should strive for an internal consistency alpha of at least .80 and a mean inter-item correlation between .15 and .50. As illustrated in Table 5, MMPI-A-RF RC scales evidenced quite variable internal consistencies (i.e, from poor (i.e., RC9 α = .47) to excellent (i.e., RCd α = .92)). Inter-item correlations were similarly variable, ranging from a disappointing .10 for RC9 to a solid .39 for RCd. Unexpectedly, only three RC scales evidenced scale homogeneity consistent with Clark and Watson's guidelines.

Importantly, Clark and Watson (1995) emphasized the importance of inter-item correlations as a very useful indicator of internal consistency. When focused exclusively on

mean inter-item correlations, all but one MMPI-A-RF RC scales meet recommended guidelines. Yet, mean inter-item correlations may mask variances in inter-item correlations. Therefore, the percentage of items falling below the .15 Clark and Watson threshold was also calculated. Unpredictably, the vast majority of RC scales evidenced at least moderate proportions of items with low correlations, with two scales—RC8 and RC9—comprising low correlations on nearly two-thirds of its items.

Table 5

Alphas and Mean Inter-item Correlations of MMPI-A-RF RC Scales

Scale	Items	α	*M* inter-item *r*	Percentage of low *r*s[a]
RCd	18	.92	.39	3.9
RC1	23	.85	.20	36.4
RC2	10	.69	.18	35.6
RC3	9	.63	.16	44.4
RC4	20	.88	.27	18.9
RC6	9	.72	.22	25.0
RC7	11	.66	.15	50.9
RC8	8	.57	.15	64.3
RC9	8	.47	.10	64.3
	M	.71	.20	38.2

Note: RCd = Demoralization; RC1 = Somatic Complaints; RC2 = Low Positive Emotions; RC3 = Cynicism; RC4 = Antisocial Behavior; RC6 = Ideas of Persecution; RC7 = Dysfunctional Negative Emotions; RC8 = Aberrant Experiences; RC9 = Hypomanic Activation. [a] Low *r*s refer to inter-item correlations < .15.

Across both measures of construct validity, RC9 performed by far the worst for scale homogeneity. Therefore, an item-level analysis was systematically conducted to see if any item removed would improve alpha. Particularly, item 182 yielded the lowest range of inter-item correlations, with several producing negative correlations. Furthermore, it produced the lowest

correlation with the overall scale (r_{it} = -.08). However, if removed, alpha only increased by .07, bringing the overall alpha to a marginal standard (α = .54). Thus, this item was retained for all study analyses given (a) the already low number of items on RC9 (i.e., 8) and (b) the minimal increase in internal consistency.

Research Question 2: Do the MMPI-A-RF RC scales evidence construct validity in a sample of youth with mental health needs?

For construct validity, MMPI-A-RF items are expected to correlate more strongly with their respective scales than other scales. As introduced in the Methods, two standards for determining convergent and discriminant validity on an item level were used to examine these correlations: Zimmerman and Mattia (2001) and Siefer et al. (2009).

Table 6

Percentage Evidencing Scaling Success and Cross-Scale Correlations for the MMPI-A-RF RC Scales

Scale	Scaling Success Criteria		Cross-Scale Correlations
	Zimmerman	Siefert	Siefert
RCd	55.6	100.0	0.0
RC1	47.8	87.0	13.0
RC2	70.0	100.0	0.0
RC3	55.6	100.0	0.0
RC4	65.0	100.0	0.0
RC6	33.3	100.0	0.0
RC7	9.1	90.9	9.1
RC8	50.0	87.5	12.5
RC9	62.5	87.5	12.5
M	49.9	94.8	5.2

Note: RCd = Demoralization; RC1 = Somatic Complaints; RC2 = Low Positive Emotions; RC3 = Cynicism; RC4 = Antisocial Behavior; RC6 = Ideas of Persecution; RC7 = Dysfunctional Negative Emotions; RC8 = Aberrant Experiences; RC9 = Hypomanic Activation. Zimmerman = Zimmerman and Mattia (2001) criteria of *r*s $\geq$ .30 for

designated scale and *r*s ≥ .20 higher for designated than non-designated scales. Siefert = Siefert et al. (2009) criteria for the proportion of item-scale comparisons with higher correlations on its respective scale than other clinical scales.

Zimmerman and Mattia (2001) specified that items must be correlated greater than .30 with its respective scale in addition to being correlated .20 higher on its respective scale versus another scale. Under this standard, 49.9% of MMPI-A-RF RC items met the criterion. In contrast to Zimmerman and Mattia's strict approach, Siefert et al. (2009) utilized a more relaxed standard. The authors defined scaling success as items evidencing a significantly greater correlation with its designated scale than non-designated clinical scales. With no-minimal difference required, Siefert et al.'s (2009) standard produced an increase in items achieving scaling success by nearly double, with 94.8% meeting criteria. Most strikingly, RC7 evidenced an over 80-point increase when applying the Siefert criteria.

Siefert et al. (2009) also described an additional approach for determining construct validity. The authors suggested calculating cross-scale correlations or the percentage of items producing an item-scale correlation with their designated scale lower than item-scale correlations with at least one other scale. Under this standard, merely 5.2% of MMPI-A-RF RC items evidenced cross-scale correlations. On a very positive note, 5 RC scales evidenced no cross-scale correlations, providing encouraging evidence of construct validity.

As an additional approach to construct validity, the current study assessed patterns of correlations between MMPI-A-RF RC scales and K-SADS-PL diagnostic sections, following previous clinical correlate research (e.g., Rogers et al., in press; Veltri et al., 2009). Broadly, patterns of convergent and discriminant validity were demonstrated for eight MMPI-A-RF RC scales. Interestingly, RC3 was the only MMPI-A-RF RC scale without demonstrable convergent validity.

It was posited that a restricted range of scores could possibly explain RC3's lack of apparent convergent validity. Yet, closer examination revealed a normal distribution of scores, with skew and kurtosis within normal limits (Skew = .64, Kurtosis = -.40). Moreover, 11 adolescents (16.7%) demonstrated clinically elevated scores. Thus, it is not clear why RC3 did not demonstrate any demonstrable convergent validity.

Table 7

Substantial Correlations (≥ .35) of MMPI-A-RF RC Scales with K-SADS-PL Diagnostic Section Scores

	K-SADS-PL Diagnostic Section										
	Internalizing						Externalizing				
	MDD	Mania	SCZ	Panic	GAD	PTSD	ODD	CD	ALC	DRG	ADHD
RCd	**.76**			.44	**.64**						
RC1	.45			.45	.47	.42					
RC2	.36				.36						
RC3											
RC4							**.56**	**.72**	**.65**	**.75**	
RC6			.38								
RC7	**.55**			.45	**.64**						
RC8			**.59**								
RC9											

Note: Large correlations (≥ .53) are bolded. RCd = Demoralization; RC1 = Somatic Complaints; RC2 = Low Positive Emotions; RC3 = Cynicism; RC4 = Antisocial Behavior; RC6 = Ideas of Persecution; RC7 = Dysfunctional Negative Emotions; RC8 = Aberrant Experiences; RC9 = Hypomanic Activation; MDD = Major Depressive Disorder; SCZ = Schizophrenia/Psychosis; GAD = Generalized Anxiety Disorder; PTSD = Posttraumatic Stress Disorder; ODD = Oppositional Defiant Disorder; CD = Conduct Disorder; ALC = Alcohol Use Disorder; DRG = Substance Use Disorder; ADHD = Attention Deficit/Hyperactivity Disorder.

Internalizing symptomatology was identified by a broad pattern of MMPI-A-RF RC scales (see Table 7). Specifically, RCd, RC1, RC2, and RC7 appeared to be most associated with K-SADS-PL disorders of an internalizing nature. Although these scales all demonstrated

correlations with internalizing symptomatology broadly, important nuances among the scales appeared.

Three K-SADS-PL diagnostic sections evaluate anxiety and trauma symptoms. Conceptually, RC7 should correlate the strongest with these sections. As a strength, RC7 evidenced moderate to large correlations with Panic Disorder and Generalized Anxiety Disorder, respectively. Surprisingly, RC7 was not the only MMPI-A-RF RC scale to correlate strongly with anxious symptomatology; RCd and RC1 also evidenced at least moderate associations with Panic Disorder and GAD. Contrary to expectations, RC7 did not evidence any substantial association with PTSD. In fact, the only MMPI-A-RF scale to correlate at least moderately with PTSD features was RC1.

Regarding depressive symptomatology, RC2 was expected to evidence large correlations with K-SADS-PL Major Depressive Disorder. In the current sample, RC2 disappointingly produced only a low moderate correlation with depressive symptoms. Evidencing poor discriminant validity, two other RC scales evidenced much stronger correlations, namely, RCd (r = .76) and RC7 (r = .55).

A pattern emerged in which several MMPI-A-RF RC scales were correlated with both depressive and anxious symptomatology. One possible explanation for such overlap is likely the comorbidity of depressive and anxious disorders in the current sample. To explore this hypothesis further, relationships among depressive and anxious symptoms was explored. Strong relationships were confirmed between (a) MDD and GAD (r = .67) and (b) MDD and Panic disorder (r = .56). In terms of diagnostic comorbidity, approximately two-thirds of adolescents diagnosed with MDD also received a GAD diagnosis (68.8%), while just over a third met for a Panic Disorder diagnosis (34.4%). Puzzlingly, the correlation between MDD and PTSD

symptoms was minimal (r = .12), despite over half of adolescents (58.1%) with MDD also being diagnosed with PTSD.

As planned, the K-SADS-PL diagnostic section specifically examining somatic symptoms was not administered in the current study. Thus, convergent and discriminant validity for RC1 could only be examined on a general internalizing/externalizing basis. From this perspective, RC1 appeared to correlate broadly with symptoms of an internalizing nature. That is, RC1 evidenced at least moderate correlations with depressive, anxious, and trauma symptoms.

Taken together, moderate to very large correlations generally support the convergent validity of RCd, RC1, RC2, and RC7. Importantly, none of these scales exhibited correlations with any externalizing symptomatology, lending strong support to their discriminant validity.

Two RC scales related to psychotic symptoms (i.e., RC6 and RC8) evidenced fairly well-defined patterns of convergent and discriminant validity. As expected, RC8 evidenced a large association (r = .59) with psychotic symptoms. Additionally, RC6 was correlated only with psychotic symptomatology, although to a low moderate degree. As noted, psychotic symptomatology is relatively infrequent among adolescents in this and other studies. Therefore, additional analyses of correlates were completed for the current study, by restricting correlates to only those adolescents exhibiting at least sub-threshold psychotic symptomatology (n = 12) on the K-SADS-PL. As expected, in a more restricted sample, RC6 increased to a large correlation (r = .60), and RC8 still remained a large correlation, decreasing slightly (r = .55).

Importantly, RC9 was also correctly correlated with manic symptomatology. Yet, the relationship only correlated to a small degree (r = .28). One possible reason for the small correlation may in part be attributed to the low representativeness of manic symptoms in the current sample. In particular, only approximately one-quarter (27.3%) described more than one

manic symptom at a clinical level.

As expected, RC4 excelled in identifying externalizing symptomatology. As a particular strength, RC4 evidenced (a) two large correlations (Oppositional Defiant Disorder and Alcohol Use Disorder) and (b) two very large correlations (Conduct Disorder and Substance Use Disorder). Furthermore, RC4 was not substantially correlated with any internalizing symptomatology. Thus, RC4 evidenced very strong convergent and discriminant validity.

MMPI-A-RF Externalizing subscales were further examined to more fully understand the relationship between RC4 and various externalizing symptomatology (see Table 8). Clearly, action-related scales provided the strongest correlates. Specifically, the two subscales centered on attitudes and beliefs (i.e., Negative School Attitudes and Antisocial Attitudes) were unrelated to any of the externalizing disorders. Closer exploration of these correlates helps to explain the differences in overall RC4 correlations between conceptually related disorders (i.e., ODD and CD; ALC and DRG). For example, DRG was strongly correlated to three externalizing subscales with a consistently very large degree (i.e., > .65), as compared to ALC's comparatively more moderate relationships with Conduct Problems and Negative Peer Influence.

Table 8

Substantial Correlations (≥ .35) of MMPI-A-RF Externalizing Subscales with K-SADS-PL Externalizing Diagnostic Section Scores

	ODD	CD	ALC	DRG	ADHD
Negative School Attitudes					
Antisocial Attitudes					
Conduct Problems	**.55**	**.68**	.51	**.65**	
Substance Abuse		**.54**	**.80**	**.72**	
Negative Peer Influence	**.58**	**.63**	.47	**.65**	
Aggression	**.57**	.40			

Note: Large correlations (≥ .53) are bolded. ODD = Oppositional Defiant Disorder; CD = Conduct Disorder; ALC = Alcohol Use Disorder; DRG = Substance Use Disorder; ADHD = Attention Deficit/Hyperactivity Disorder.

Research Question 3: Do the MMPI-A-RF RC scales accurately classify relevant DSM-5 diagnoses?

While not explicitly intended to be a diagnostic instrument, clinical inferences are commonly drawn from multiscale inventories, such as the MMPI-A-RF. Archer (2017), one of the authors of the MMPI-A-RF, cautioned against using it to determine single diagnoses; instead, he emphasized the MMPI-A-RF's value of generating different diagnostic possibilities. Such emphasis is evidenced by the inclusion of interpretive statements and diagnostic considerations for clinical elevations on each RC scale (Archer et al., 2016). As such, it is a critically important task to explore the accuracy of such diagnostic inferences. Research Question 3 addressed this task directly.

DSM-5 diagnoses for Research Question 3 were derived from item responses on the K-SADS-PL. Because the K-SADS-PL was published prior to DSM-5, additional questions targeting new criteria were developed. For example, the following question was added to determine whether any significant maladaptive changes to an adolescent's life were made to avoid future panic attacks: "Have these attacks caused you to change your life in major ways? Have you gone out of your way to prevent these attacks from happening again?" In total, 11 DSM-5 diagnoses were examined: Major Depressive Disorder (MDD), Bipolar Disorder, Schizophrenia, Panic Disorder, Generalized Anxiety Disorder (GAD), Attention Deficit Hyperactivity Disorder (ADHD), Oppositional Defiant Disorder (ODD), Conduct Disorder (CD), Alcohol Use Disorder, Substance Use Disorder (SUD), and Posttraumatic Stress Disorder (PTSD). Following previous clinical correlate research (Cashel et al., 1998), only those diagnostic categories with sufficient frequencies (i.e., more than ten persons meeting diagnostic threshold) were considered for Research Question 3.

The RC7 scale emerged as the most effective MMPI-A-RF RC scale in predicting internalizing diagnoses, specifically MDD and GAD. In both instances, as scores on RC7 increased, so did the odd ratios of receiving the specific diagnosis (see Table A.1 in the Appendix). Furthermore, RC7 yielded large effect sizes (MDD $d = 1.19$ and GAD $d = 1.03$) in discriminating between adolescents meeting criteria for either MDD or GAD and adolescents who were sub-threshold (see Table A.1 in the Appendix). Thus, elevated RC7 scores can provide clinically relevant information. Yet, it is also important to consider the utility of such information.

Table 9

Effectiveness of MMPI-A-RF RC7 Cut Scores (RC7 $\geq$ 60) for Identifying Major Depressive Disorder and Generalized Anxiety Disorder Diagnoses

Diagnosis	Sensitivity	Specificity	OCC	Base Rate [a]	
				PPP	NPP
Major Depressive Disorder	.34	.91	.64	.33	.92
Generalized Anxiety Disorder	.35	.91	.60	.11	.98

Note: OCC = overall correct classification; PPP = positive predictive power; NPP = negative predictive power. [a] Following recent adolescent epidemiological research, an 11% base rate was used for Major Depressive Disorder (Avenevoli, Swendsen, He, Burstein, & Merikangas, 2015), while a 3% base rate was used for Generalized Anxiety Disorder (Burstein, Beesdo-Baum, He, & Merikangas, 2014).

Utility estimates (i.e., sensitivity and specificity) were calculated to determine how useful elevated RC7 scores were in determining MDD and GAD diagnoses. As summarized in Table 9, RC7 threshold cut scores performed similarly across MDD and GAD diagnoses. In both instances, elevated RC7 scores evidenced rather low sensitivity (ranging from .34 to .35), but excellent specificity (.91). Such poor sensitivities militate against the use of the RC7 as a good indicator to rule in either MDD and/or GAD. Encouragingly, high specificities coupled with excellent negative predictive power statistics lend support for the use of RC7 cut scores to rule-

out MDD and/or GAD. To put it simply, adolescents producing RC7 scores less than 60 are unlikely to meet criteria for MDD and/or GAD diagnoses. In practice, this is valuable information for clinicians as they can be confident that individuals with low RC7 scores are not experiencing clinically significant depressive or anxious symptoms.

In the externalizing realm, RC4 was particularly effective at predicting conduct disorders and substance use disorders. Specifically, as RC4 *T* scores increased, the odd ratios of receiving a Conduct Disorder or Substance Use Disorder diagnosis increased marginally. Moreover, RC4 effectively discriminated between adolescents meeting criteria for CD (d = 1.98) and SUD (d = 1.99) and those not meeting criteria for such disorders, to a very large degree (see Table A.1 in the appendix).

Utility estimates must be examined in order to understand the clinical implications of elevated RC4 scores. Importantly, the MMPI-A-RF manual (Archer et al., 2016) specifies two levels of interpretation for RC4: clinically elevated with *T* scores greater than or equal to 60 and *T* scores greater than or equal to 80. The effectiveness of RC4 in identifying CD and SUD was calculated at both levels. Encouragingly, an RC4 cut score $\geq$ 60 achieved well-balanced utility estimates in identifying CD. That is, it maintained adequate sensitivity, accurately identifying nearly three quarters of the youth meeting criteria for a CD diagnosis, while still producing an excellent specificity. The same cut score also produced excellent specificity when applied to substance use disorder; however, it resulted in a predictable drop in sensitivity. As expected, raising the cut score $\geq$ 80 produced near-perfect and perfect specificity for CD and SUD, respectively, but slashed the sensitivity by approximately half. Overall, these results suggest that an RC4 cut score $\geq$ 60 can be effective as a general indicator in ruling out CD and/or SUD.

Further, using the higher RC4 cut score $\geq$ 80 can increase clinicians' confidence in ruling out such diagnoses, but at the cost of missing some adolescents exhibiting significant symptoms.

Table 10

Effectiveness of MMPI-A-RF RC4 Cut Scores for Identifying Conduct Disorder and Substance Use Disorder Diagnoses

Diagnosis	Sensitivity	Specificity	OCC	Base Rate[a]	
				PPP	NPP
		RC4 $\geq$ 60			
Conduct Disorder	.73	.92	.88	.40	.98
Substance Use Disorder	.59	.95	.85	.56	.96
		RC4 $\geq$ 80			
Conduct Disorder	.27	.96	.80	.33	.95
Substance Use Disorder	.29	1.00	.80	1.00	.93

Note: OCC = overall correct classification; PPP = positive predictive power; NPP = negative predictive power. [a] Following recent adolescent epidemiological research (Merikangas et al., 2010), a 7% base rate was used for Conduct Disorder, while a 9% base rate was used for Substance Use Disorder.

Research Question 4: Are MMPI-A-RF RC scales' interpretive statements related to serious mental disorders?

Item-level correlates provide the most nuanced appraisal of RC scale interpretive statements. Thus, Research Question 4 examined how MMPI-A-RF RC scales correlated with specific K-SADS-PL symptomatology. Item-level correlates were divided broadly into internalizing and externalizing domains for ease of interpretation.

Three MMPI-A-RF RC scales (i.e., RCd, RC1, and RC7) shared substantial numbers of internalizing-related correlates. More specifically, each of the three scales exhibited moderate to large correlations with depressive and anxious features (see Table 11). Yet, important differences emerged upon additional analysis.

Table 11

Substantial K-SADS-PL Internalizing Item-Level Correlates (≥ .35) for the MMPI-A-RF RC Scales

K-SADS-PL		MMPI-A-RF RC Scales								
Section	K-SADS-PL Item Stem	RCd	RC1	RC2	RC3	RC4	RC6	RC7	RC8	RC9
MDD	Depressed mood	**.65**	.43					.41		
	Irritability and anger									
	Anhedonia	**.64**		.39						
	Recurrent thoughts of death	**.61**	.35					.36		
	Suicidal ideation	**.53**						.36		
	Self harm	.46								
	Lack of reactivity	**.68**	.45	.39				.41		
	Quality of dysphoria	.50								
	Sleep disturbance - initial		.36					.35		
	Sleep disturbance - middle									
	Sleep disturbance - terminal									
	Non-restorative sleep	.38	.39							
	Sleep disturbance - hypersomnia									
	Fatigue	.51	.42							
	Disturbance in concentration						.36			
	Indecision									
	Decreased Appetite	.45						.38		
	Weight loss							.36		
	Increased Appetite									
	Weight gain									

K-SADS-PL		MMPI-A-RF RC Scales								
Section	K-SADS-PL Item Stem	RCd	RC1	RC2	RC3	RC4	RC6	RC7	RC8	RC9
	Psychomotor agitation									
	Psychomotor retardation									
	Worthlessness	.45								
	Guilt									
	Hopelessness	.49		.40						
Mania	Elevated mood									
	Decreased need for sleep									
	Increased goal-directed activity									
	Racing thoughts									
	Grandiosity									
	Pressured speech									
	Poor judgment					.48				
	Distractibility									
	Physical restlessness									
	Influence of alcohol or drugs					.41				
Psychosis	Hallucinations						.35		**.61**	
	Delusions									
	Flat affect								.46	
	Inappropriate affect						.38		**.53**	
	Incoherence						.37		**.57**	
	Loosening of associations						.37		**.57**	
	Catatonic behavior						.37		**.57**	

K-SADS-PL		MMPI-A-RF RC Scales								
Section	K-SADS-PL Item Stem	RCd	RC1	RC2	RC3	RC4	RC6	RC7	RC8	RC9
Panic Disorder	Panic attacks	.47	.48					.42		
	Fear of another attacks	.47	.38					.42		
	Agoraphobia	.49	.42					.45		
	Maladaptive changes to avoid attack							.39		
GAD	Excessive worry	.44						.42		
	Overconcern with competence							.40		
	Need for reassurance							.36		
	Ability to control worries	.41						.41		
	Muscle tension	.48	.47					.41		
	Restlessness	.38	.35					.51		
	Easily fatigued	.39	.35					.42		
	Feeling keyed up	.37								
	Difficulty concentrating	.43								
	Trouble falling or staying asleep	.35								
	Irritability									
	Recurrent thoughts or images									
PTSD	Efforts to avoid thoughts or feelings									
	Nightmares					.40				
	Insomnia					.41				
	Irritability					.35				
	Re-enactment									
	Dissociative episodes									
	Distress to exposure									

K-SADS-PL		MMPI-A-RF RC Scales								
Section	K-SADS-PL Item Stem	RCd	RC1	RC2	RC3	RC4	RC6	RC7	RC8	RC9
	Efforts to avoid physical reminders									
	Inability to recall details									
	Diminished interest in activities									
	Detachment					.42				
	Restricted affect									
	Sense of foreshortened future									
	Difficulty concentrating									
	Hypervigilance					.37				
	Exaggerated startle response									
	Physiological reactivity									
	Flashbacks									
	Negative beliefs									
	Distorted cognitions									
	Negative emotional state									
	Inability to experience positive emotions									
	Reckless or self-destructive behavior					.48				

Note: Large correlations ($\geq$ .53) are bolded. RCd = Demoralization; RC1 = Somatic Complaints; RC2 = Low Positive Emotions; RC3 = Cynicism; RC4 = Antisocial Behavior; RC6 = Ideas of Persecution; RC7 = Dysfunctional Negative Emotions; RC8 = Aberrant Experiences; RC9 = Hypomanic Activation; MDD = Major Depressive Disorder; GAD = Generalized Anxiety Disorder; PTSD = Posttraumatic Stress Disorder.

In general, RCd evidenced the most robust relationships among affectively related correlates. It produced very large associations with characteristic depressive symptomatology, including depressed mood (r = .65), anhedonia (r = .64), and recurrent thoughts of death (r = .61). Furthermore, it demonstrated moderate to large correlations with nine other MDD correlates. Of note, RCd also produced several unique item-level correlates, such as non-suicidal self-harm (r = .46) and worthlessness (r = .45). Yet, neither RCd, nor any other MMPI-A-RF RC scale, produced significant correlations with manic symptomatology.

In stark contrast, RC2 produced only three empirical correlates with MDD. A surprising pattern developed within these three empirical correlates in that each correlate was shared with at least one other RC scale; namely, RCd. Moreover, these features tended to be correlated to a lesser degree with RC2 than RCd. As a persuasive example, anhedonia was considerably much more correlated with RCd (r = .64) than with RC2 (r = .39).

With respect to anxiety, RCd and RC7 evidenced numerous item-level correlates with anxiety-relevant K-SADS-PL diagnostic sections. More specifically, RCd and RC7 demonstrated at least moderate correlations with eleven correlates each among Panic Disorder and Generalized Anxiety Disorder correlates. Interestingly, RC7 was uniquely related to the maladaptive changes made to avoid future panic attacks (r = .39), overconcern with competence (r = .40), and an excessive need for reassurance (r = .36). Meanwhile, RCd was uniquely correlated with three consequences of excessive worry: feeling keyed up (r = .37), difficulty concentrating (r = .43), and difficulty sleeping (r = .35). Thus, it appears the two scales actually measure different aspects of anxiety, with RC7 related to apprehension and RCd more so related to its manifestation.

Most surprising, PTSD demonstrated a drastically different pattern of correlations than the anxiety disorders. While RCd and RC7 were correlated with most aspects of Panic Disorder and GAD, neither scale was significantly correlated with PTSD. In fact, the majority of PTSD features were at least moderately correlated with RC4. As expected, acting out behavior as a result of trauma was most strongly correlated with RC4; reckless or self-destructive behavior produced the largest association with RC4 (r = .48).

To explore this unanticipated relationship further, item-level correlates were reexamined in a sample restricted to only those adolescents meeting criteria for a PTSD diagnosis (n = 28). In this more limited sample, the only items remaining correlated with RC4 were detachment and reckless or self-destructive behavior (see Table 12). Instead, the PTSD correlates in the more restricted sample reverted to a similar pattern as the anxiety disorders. That is, RCd emerged as the most relevant scale, with RC7 also contributing some correlates. Thus, initially including non-clinically threshold PTSD symptoms may have obscured these relationships.

Of interest, RC3 evidenced a relationship with several PTSD symptoms in the more refined sample. Specifically, four moderate correlations and one large correlation (i.e., distorted cognitions, r = .53) surfaced. Upon additional consideration, it follows plausibly that those suffering from PTSD would likely experience some level of cynicism and interpersonal distrust, particularly if their reported trauma was interpersonal in nature. In the current sample, over 80% of those meeting criteria for PTSD experienced at least one interpersonal trauma (e.g., physical or sexual abuse).

Despite their relative infrequence in the current sample (M = 0.24, SD = 0.41), moderate to large correlations were still exhibited among psychotic symptoms. As expected from a conceptual standpoint, psychotic features and negative symptoms showed a moderate to large

relationship with RC6 and RC8. Upon further examination, RC6 and RC8 were virtually unrelated to the vast majority of internalizing symptomatology, with the exception of difficulty concentrating (see Table 13).

As with the scale-level correlates, RC4 consistently showed the strongest relationship with externalizing symptomatology. Further examination revealed moderate to very large correlations with ODD (*M r* = .47), CD (*M r* = .55), Alcohol Use Disorder (*M r* = .61), and Substance Use Disorder (*M r* = .68). Virtually no significant correlations were observed between RC4, or any other MMPI-A-RF RC scales, and ADHD symptomatology.

As evidence of discriminant validity, externalizing symptomatology produced multiple negative correlations with scales previously discussed as linked to internalizing symptomatology. Examples include RCd (*M* = -.39) and RC2 (*M* = -.38) scales. These findings have important clinical implications for RCd. RCd is described as the common underlying distress or demoralization factor across clinical disorders; however, the production of negative correlations with externalizing symptomatology demonstrates the distress or demoralization is focused specifically on internalized or affective distress.

Given RC4's strong relationship with externalizing symptomatology broadly, it is critical to also investigate how externalizing subscales contribute to interpretive statements. To this end, the various relationships between externalizing MMPI-A-RF subscales and K-SADS-PL externalizing items were explored. Overall, Conduct Problems (CNP), Substance Abuse (SUB), Negative Peer Influence (NPI), and Aggression (AGG) produced numerous item-level correlates that were diagnostically relevant.

Table 12

Substantial K-SADS-PL PTSD Item-Level Correlates (≥ .35) for the MMPI-A-RF RC Scales in a Sample of Adolescents Meeting Criteria for PTSD

K-SADS-PL Item Stem: PTSD	MMPI-A-RF RC Scales								
	RCd	RC1	RC2	RC3	RC4	RC6	RC7	RC8	RC9
Recurrent thoughts or images									
Efforts to avoid thoughts or feelings									
Nightmares				.37					
Insomnia									
Irritability									
Re-enactment									
Dissociative episodes				.43					
Hallucinations								.51	
Distress to exposure									
Efforts to avoid physical reminders	.39			.45					
Inability to recall details			.40						
Diminished interest in activities									
Detachment					.35				
Restricted affect									
Sense of foreshortened future									
Difficulty concentrating	.46								

K-SADS-PL Item Stem: PTSD	MMPI-A-RF RC Scales								
	RCd	RC1	RC2	RC3	RC4	RC6	RC7	RC8	RC9
Hypervigilance									
Exaggerated startle response									
Physiological reactivity	.41	.44		.43			.47		
Flashbacks									
Negative beliefs			-.34						
Distorted cognitions	**.64**			**.53**			.47		.42
Negative emotional state	.47							-.37	
Inability to experience positive emotions	.52						.41		
Reckless or self-destructive behavior					.45				

Note: Large correlations ($\geq$ .53) are bolded. RCd = Demoralization; RC1 = Somatic Complaints; RC2 = Low Positive Emotions; RC3 = Cynicism; RC4 = Antisocial Behavior; RC6 = Ideas of Persecution; RC7 = Dysfunctional Negative Emotions; RC8 = Aberrant Experiences; RC9 = Hypomanic Activation; PTSD = Posttraumatic Stress Disorder.

Table 13

Substantial K-SADS-PL Externalizing Item-Level Correlates (≥ .35) for the MMPI-A-RF RC Scales

K-SADS-PL		MMPI-A-RF RC Scales								
Section	K-SADS-PL Item Stem	RCd	RC1	RC2	RC3	RC4	RC6	RC7	RC8	RC9
ADHD	Difficulty sustaining attention									
	Easily distracted									
	Difficulty remaining seated									
	Impulsivity			-.36						
	Careless mistakes									
	Doesn't listen									
	Difficulty following instructions									
	Difficulty organizing tasks									
	Dislikes tasks requiring attention									
	Loses things									
	Forgetful									
	Fidget									
	Runs or climbs									
	On the go									.35
	Difficulty playing quietly									
	Blurts out answers									
	Difficulty waiting turn									
	Interrupts or intrudes									
	Shifts activities									
	Talks excessively									

K-SADS-PL		MMPI-A-RF RC Scales								
Section	K-SADS-PL Item Stem	RCd	RC1	RC2	RC3	RC4	RC6	RC7	RC8	RC9
	Engages in dangerous activities					.39				
ODD	Loses temper					.40	.42			
	Argues a lot with adults									
	Disobeys rules					**.65**				
	Easily annoyed					.37				
	Angry or resentful									
	Spiteful or vindictive					.47	.47			
	Uses bad language	-.40		-.39		.49				
	Annoys people on purpose					.46				
	Blames others for mistakes									
Conduct Disorder	Lies	-.35				**.53**				
	Truant					**.55**				
	Initiates physical fights	-.42				**.58**				
	Bullies, threatens, or intimidates					.37	.42			
	Nonaggressive stealing					**.59**				
	Vandalism					**.58**				
	Breaking and entering					.48				
	Aggressive stealing					**.55**				
	Fire setting									
	Often stays out at night					**.60**				
	Ran away overnight					**.67**				
	Use of a weapon					**.53**				
	Physical cruelty to others					.52				

K-SADS-PL		MMPI-A-RF RC Scales								
Section	K-SADS-PL Item Stem	RCd	RC1	RC2	RC3	RC4	RC6	RC7	RC8	RC9
	Forced sexual activity									
	Cruelty to animals									
Alcohol Use	Frequency					**.55**				
	Concern from others					**.59**				
	Drinks more than planned					**.53**				
	Negative physical consequences					**.57**				
	Dangerous behavior					**.64**				
	Negative psych. consequences					**.61**				
	Negative occupational consequences					**.66**				
	Negative social consequences					**.68**				
	Legal consequences					**.57**				
	Failure to fulfill major obligations					**.58**				
	Important activities given up					**.74**				
	Time consuming					**.60**				
	Tolerance					**.59**				
	Tried to quit					**.60**				
	Withdrawal					**.60**				
	Alcohol used to relieve withdrawal				.36	**.61**				
Substance Use	Frequency					**.72**				
	Uses more than planned					**.60**				
	Negative physical consequences					**.68**				
	Dangerous behavior					**.76**				
	Negative psych. consequences					**.63**				

K-SADS-PL		MMPI-A-RF RC Scales								
Section	K-SADS-PL Item Stem	RCd	RC1	RC2	RC3	RC4	RC6	RC7	RC8	RC9
	Negative occupational consequences					**.75**				
	Negative social consequences					**.66**				
	Legal consequences	-.39				**.75**				
	Failure to fulfill major obligations					**.68**				
	Important activities given up					**.71**				
	Time consuming					**.69**				
	Tolerance					**.70**				
	Tried to quit					**.61**				
	Withdrawal					**.66**				
	Drugs used to relieve withdrawal					**.67**				

Note: Large correlations ($\geq$ .53) are bolded. RCd = Demoralization; RC1 = Somatic Complaints; RC2 = Low Positive Emotions; RC3 = Cynicism; RC4 = Antisocial Behavior; RC6 = Ideas of Persecution; RC7 = Dysfunctional Negative Emotions; RC8 = Aberrant Experiences; RC9 = Hypomanic Activation. ODD = Oppositional Defiant Disorder; ADHD = Attention Deficit/Hyperactivity Disorder.

In general, MMPI-A-RF subscales evidenced strong relationships on conceptually relevant items. For example, SUB demonstrated the largest associations with alcohol use disorders (*M r* = .79) and SUD (*M r* = .66). More specifically, several items were comparatively more related to SUB than all other externalizing subscales. In fact, many items produced very large associations, with more than 10 items exceeding an *r* of .70. Such statistics provide compelling evidence of convergent validity.

As expected, CNP was clearly most related to CD symptoms. In several instances, CD items were relatively more strongly related to CNP than its externalizing subscale counterparts. In particular, initiating physical fights (*r* = .62), often staying out at night (*r* = .62), and using a weapon (*r* = .53) all of which produced large associations with CNP but comparatively smaller relationships with other subscales. Additionally, AGG was most strongly correlated with aggressive attitudes or actions, such as (a) being spiteful or vindictive (*r* = .54) or (b) bullying, threatening, or intimidating others (*r* = .52).

Table 14

Substantial K-SADS-PL Externalizing Item-Level Correlates (≥ .35) for the MMPI-A-RF Externalizing Subscales

K-SADS-PL		Externalizing Subscale					
Section	K-SADS-PL Item Stem	NSA	ASA	CNP	SUB	NPI	AGG
ADHD	Difficulty sustaining attention						
	Easily distracted						
	Difficulty remaining seated						
	Impulsivity			.41			
	Careless mistakes						
	Doesn't listen						
	Difficulty following instructions						
	Difficulty organizing tasks						
	Dislikes tasks requiring attention						

K-SADS-PL		Externalizing Subscale					
Section	K-SADS-PL Item Stem	NSA	ASA	CNP	SUB	NPI	AGG
	Loses things						
	Forgetful						
	Fidget						
	Runs or climbs						
	On the go						
	Difficulty playing quietly						
	Blurts out answers						
	Difficulty waiting turn						
	Interrupts or intrudes						
	Shifts activities						
	Talks excessively						
	Engages in dangerous activities			.42		.46	
ODD	Loses temper			.48		.35	.41
	Argues a lot with adults			.35		.46	.42
	Disobeys rules			**.60**	.51	**.62**	.35
	Easily annoyed					.40	.41
	Angry or resentful						.39
	Spiteful or vindictive			.50		.40	**.54**
	Uses bad language			.48	.35	.42	
	Annoys people on purpose			.38		**.56**	.45
	Blames others for mistakes						.42
Conduct Disorder	Lies			.49	.39	.49	.30
	Truant			.43	.44	.50	
	Initiates physical fights			**.62**	.35	.52	.43
	Bullies, threatens, or intimidates			.41		.36	.52
	Nonaggressive stealing			**.56**	**.57**	.45	.36
	Vandalism			**.61**	.41	.44	
	Breaking and entering			.43	.40	.44	
	Aggressive stealing			.48	.49	.46	
	Firesetting						
	Often stays out at night			**.62**	.46	.48	
	Ran away overnight			**.55**	**.60**	**.55**	

K-SADS-PL		Externalizing Subscale					
Section	K-SADS-PL Item Stem	NSA	ASA	CNP	SUB	NPI	AGG
	Use of a weapon			**.53**	.35	.43	.36
	Physical cruelty to others			.49	.39	.48	.46
	Forced sexual activity						
	Cruelty to animals						
Alcohol Use	Frequency			.46	**.67**	.45	
	Concern from others			.47	**.69**	.45	
	Drinks more than planned			.38	**.65**	.41	
	Negative physical consequences			.43	**.78**	.38	
	Dangerous behavior			.50	**.75**	.48	
	Negative psych. consequences			.45	**.77**	.45	
	Negative occupational consequences			.55	**.80**	.41	
	Negative social consequences			.51	**.83**	.48	
	Legal consequences			.44	**.71**	.40	
	Failure to fulfill major obligations			.47	**.70**	.42	
	Important activities given up			.57	**.83**	**.55**	
	Time consuming			.47	**.78**	.41	
	Tolerance			.46	**.71**	.45	
	Tried to quit			.47	**.76**	.41	
	Withdrawal			.48	**.75**	.41	
	Alcohol consumed to relieve withdrawal			.47	**.73**	.44	
Substance Use	Frequency			**.59**	**.74**	**.62**	
	Uses more than planned			**.53**	**.60**	**.54**	
	Negative physical consequences			**.66**	**.54**	**.61**	
	Dangerous behavior			**.63**	**.70**	**.69**	
	Negative psych. consequences			**.54**	**.59**	**.56**	
	Negative occupational consequences			**.64**	**.73**	**.65**	
	Negative social consequences			**.61**	**.64**	**.56**	
	Legal consequences			**.66**	**.69**	**.63**	
	Failure to fulfill major obligations			**.60**	**.62**	**.58**	
	Important activities given up			**.62**	**.68**	**.59**	
	Time consuming			**.57**	**.75**	**.58**	
	Tolerance			**.54**	**.76**	**.61**	

K-SADS-PL		Externalizing Subscale					
Section	K-SADS-PL Item Stem	NSA	ASA	CNP	SUB	NPI	AGG
	Tried to quit			**.56**	**.57**	**.54**	
	Withdrawal			**.59**	**.65**	**.57**	
	Drugs consumed to relieve withdrawal			**.62**	**.64**	**.56**	

Note. Large correlations (≥ .53) are bolded. NSA = Negative School Attitudes, ASA = Antisocial Attitudes, CNP = Conduct Problems, SUB = Substance Abuse, NPI = Negative Peer Influence, AGG = Aggression; ODD = Oppositional Defiant Disorder; ADHD = Attention Deficit/Hyperactivity Disorder.

Research Question 5: Do non-gendered norms produce accurate clinical interpretations?

Given the gendered versus non-gendered norms debate outlined in the Introduction, Research Question 5 sought to examine how gender impacts the accuracy of clinical interpretations. To fully investigate this issue, gender differences were examined in terms of MMPI-A-RF RC scale elevations, in addition to the subsequent clinical correlates.

Statistically significant gender differences were pronounced on several MMPI-A-RF RC scales (see Table 15). A series of one-way analyses of variance (ANOVAs) revealed that males and females produced different levels of elevations on RCd, RC1, RC4, and RC7. When examining these differences more closely, an interesting pattern emerged in which females were elevated significantly higher on internalizing-relevant scales (i.e., RCd, RC1, and RC7), while their male counterparts were significantly more elevated on RC4, an externalizing-relevant scale.

Clinical correlates were examined by gender to explore how such gender differences affected MMPI-A-RF RC scale interpretability. Specifically, correlations between MMPI-A-RF RC scales and K-SADS-PL total scores for each diagnostic section were reviewed for each gender. While some scales appeared to function similarly across genders, some specific gender differences appeared (see Table 16).

Table 15

Differences in MMPI-A-RF RC Scale Elevations between Females and Males

MMPI-A-RF	Total sample (N = 66)		Gender				F	p	d
			Females (n = 40)		Males (n = 26)				
Scale	*M*	*SD*	*M*	*SD*	*M*	*SD*			
RCd	61.14	15.03	66.65	13.41	52.65	13.53	17.04	**<.001**	1.04
RC1	58.74	13.90	61.60	13.93	54.35	12.90	4.53	**.04**	0.54
RC2	56.36	12.29	57.60	12.18	54.46	12.44	1.03	.31	0.26
RC3	53.48	11.12	52.40	11.24	55.15	10.94	0.97	.33	-0.25
RC4	53.38	13.85	50.38	13.14	58.00	13.88	5.08	**.03**	-0.56
RC6	52.73	11.53	51.45	9.75	54.69	13.80	1.25	.27	-0.28
RC7	52.24	10.89	55.18	10.36	47.73	10.29	8.18	**.01**	0.72
RC8	51.48	9.99	51.58	9.94	51.35	10.26	0.01	.93	0.02
RC9	45.91	8.55	45.95	8.05	45.85	9.44	0.01	.96	0.01

Note: RCd = Demoralization; RC1 = Somatic Complaints; RC2 = Low Positive Emotions; RC3 = Cynicism; RC4 = Antisocial Behavior; RC6 = Ideas of Persecution; RC7 = Dysfunctional Negative Emotions; RC8 = Aberrant Experiences; RC9 = Hypomanic Activation.

Encouragingly, approximately half of the RC scales produced similar clinical correlates across genders. Two of the RC scales (i.e., RC3 and RC9) did not produce any substantial (i.e., $\geq$ .35) correlations with K-SADS-PL diagnostic sections. The remaining three RC scales (i.e., RCd, RC4, and RC7) revealed fairly consistent patterns of clinical correlates. That is, the scales evidenced substantial correlations on the majority of the same K-SADS-PL diagnostic sections yielded only small differences (i.e., $\leq$ .10). As an example, females (M = .71) exhibited slightly stronger correlations with RC4 than males (M = .61); although, both correlations were very large in nature. As an exception, males also produced moderate correlations with manic symptoms on RC4; no such associations were exhibited in females.

Table 16

Substantial Correlations (≥ .35) of MMPI-A-RF RC Scales by Gender with the K-SADS-PL Diagnostic Section Total Scores

MMPI-A-RF Scales	Gender	K-SADS-PL Diagnostic Section										
		Internalizing						Externalizing				
		MDD	Mania	SCZ	Panic	GAD	PTSD	ODD	CD	ALC	DRG	ADHD
RCd	M	**.67**				**.53**						
	F	**.70**				.51						
RC1	M	**.55**				.46				-.36		.58
	F				.43	.39						
RC2	M		-.41						-.43	-.50		
	F	.36				.37						
RC3	M											
	F											
RC4	M		.50				.40	.42	**.72**	**.62**	**.69**	
	F						.37	**.61**	**.71**	**.74**	**.78**	
RC6	M			.54			.36	**.56**				
	F					.41				.36		
RC7	M	.56				**.59**		.37				.41
	F	.41				**.59**						
RC8	M			**.79**			**.53**	.50				
	F	.35		.43	.37							
RC9	M											
	F											

Note: Large correlations (≥ .53) are bolded. RCd = Demoralization; RC1 = Somatic Complaints; RC2 = Low Positive Emotions; RC3 = Cynicism; RC4 = Antisocial Behavior; RC6 = Ideas of Persecution; RC7 = Dysfunctional Negative Emotions; RC8 = Aberrant Experiences; RC9 = Hypomanic Activation; MDD = Major Depressive Disorder; SCZ = Schizophrenia/Psychosis; GAD = Generalized Anxiety Disorder; PTSD = Posttraumatic Stress Disorder; ODD = Oppositional Defiant Disorder; CD = Conduct Disorder; ALC = Alcohol Use Disorder; DRG = Substance Use Disorder; ADHD = Attention Deficit/Hyperactivity Disorder; M = Male (n = 26); F = Female (n = 40).

Contrastingly, the remaining four RC scales exhibited drastically different clinical correlates when separated by gender, potentially limiting the generalizability across genders. To illustrate, the most striking gender differences occurred on scales RC6 and RC8. While previous analyses primarily indicated RC6 and RC8 to be related to psychotic symptomatology (see Research Question 2), only males appeared to follow this pattern. To more fully understand this anomaly, item-level correlates for each scale, separated by gender, were examined. Analyses revealed males exhibited moderate to very large correlations (*r*s ranging from .38 to .81) for each psychotic symptom across both scales. Conversely, females exhibited only moderate correlations (*r*s ranging from .43 to .49) on RC8; RC6 did not produce any substantial correlations. Such poor convergent validity in female adolescents possibly militates against the scales use in this population. It was posited that one possible explanation for the poor convergent validity could result from a potentially restricted range of scores in females. Despite a skewed distribution, distribution of scores in females closely mirrored the distribution of scores in males. Thus, it is not apparently clear why the scales perform so differently for females.

Research Question 6: Are MMPI-A-RF Higher-Order scales' interpretive statements related to serious mental disorders?

Following the structure of the MMPI-2-RF, the authors of the MMPI-A-RF developed three broad higher-order (H-O) scales: Emotional/Internalizing Dysfunction (EID), Thought Dysfunction (THD), and Behavioral/Externalizing (BXD). According to Archer et al. (2016), clinicians should first review elevations on H-O scales for interpretation before evaluating RC scales. Thus, a thorough examination of the H-O scales was deemed essential.

As an initial step, patterns between H-O scales and broad psychopathology diagnostic categories were assessed. As described in Table 17, H-O scales were clearly related to diagnostic

categories most conceptually related to their namesake, yielding good convergent validity. EID evidenced its largest associations with internalizing disorders, including MDD and GAD. Encouragingly, THD was largely related to psychotic symptomatology, with no other substantial correlations. Furthermore, BXD was strongly related to four externalizing disorders, with the highest correlation for CD (r = .71) and the other three ranging from .58 to .64. As compelling evidence of discriminant validity, none of the H-O scales shared any overlapping correlations with any other H-O scale. Thus, the H-O scales can be considered psychometrically sound.

Table 17

Substantial Correlations (≥ .35) of MMPI-A-RF H-O scales with K-SADS-PL Diagnostic Section Scores

	K-SADS-PL Diagnostic Section										
	Internalizing						Externalizing				
	MDD	Mania	SCZ	Panic	GAD	PTSD	ODD	CD	ALC	DRG	ADHD
EID	**.78**			.47	**.70**						
THD			**.55**								
BXD							**.58**	**.71**	**.62**	**.64**	

Note: Large correlations (≥ .53) are bolded. EID = Emotional/Internalizing Dysfunction; THD = Thought Dysfunction; BXD = Behavioral/Externalizing Dysfunction. MDD = Major Depressive Disorder; SCZ = Schizophrenia/Psychosis; GAD = Generalized Anxiety Disorder; PTSD = Posttraumatic Stress Disorder; ODD = Oppositional Defiant Disorder; CD = Conduct Disorder; ALC = Alcohol Use Disorder; DRG = Substance Use Disorder; ADHD = Attention Deficit/Hyperactivity Disorder.

In clinical practice, the H-O scales represent the first stage of the interpretive process, followed by the more refined interpretation of the RC scales. As a result, it is critical to investigate the interpretive value of the H-O scales. To fully assess their potential utility, item-level clinical correlates were assessed between H-O scales and K-SADS-PL items, divided into internalizing and externalizing domains (see Tables 18 and 19). Generally, the H-O clinical correlates closely mirrored those of the RC scales, clustering consistently with the correct

corresponding scales. Yet, important differences between the levels of correlates did emerge in terms of magnitude and coverage.

Differences between H-O and RC scales in the strength of the clinical correlates varied considerably. On one end of the spectrum, EID produced slightly stronger correlations with internalizing-relevant content on both a scale and item level than its related RC scales. Such differences in magnitude were particularly apparent when comparing EID and RC1 or RC2 correlates. However, on the other end of the spectrum, the magnitude of BXD correlates with externalizing-relevant content was consistently lower than that of its RC4 counterpart. In particular, a pattern emerged in which BXD produced consistently lower correlations with substance abuse (including alcohol abuse) content as compared to RC4. To illustrate, the strength of the correlation with the K-SADS-PL Substance Abuse diagnostic section dropped by over one-tenth when looking at the broader BXD scale (BXD r = .64 vs. RC4 r = .75). THD represented the middle of the spectrum by producing consistently stronger correlations than RC6, but weaker correlations than RC8.

As expected, when utilizing the broader H-O scales, some specificity in terms of clinical correlates was lost. A pattern developed in which clinical correlates related to sleep disturbance (e.g., trouble falling or staying asleep) were not identified by EID, despite producing moderate elevations with at least one of its corresponding RC scales. As another pattern, several PTSD features previously identified by RC4 (i.e., irritability and hypervigilance) were not subsequently identified by BXD. Notably, as a trade off, several items previously unidentified by RC scales emerged when looking more broadly. For example, EID was moderately correlated with weight loss and racing thoughts, meanwhile BXD was uniquely related to impulsivity and arguing a lot with adults.

Table 18

Substantial K-SADS-PL Internalizing Item-Level Correlates (≥ .35) for the MMPI-A-RF H-O Scales

K-SADS-PL		MMPI-A-RF H-O Scales		
Section	K-SADS-PL item stem	EID	THD	BXD
MDD	Depressed mood	**.69**		
	Irritability and anger			
	Anhedonia	**.63**		
	Recurrent thoughts of death	**.62**		
	Suicidal ideation	**.53**		
	Self harm	.47		
	Lack of reactivity	**.70**		
	Quality of dysphoria	.43		
	Sleep disturbance - initial			
	Sleep disturbance - middle			
	Sleep disturbance - terminal			
	Non-restorative sleep			
	Sleep disturbance - hypersomnia			
	Fatigue	**.53**		
	Disturbance in concentration	.35	.36	
	Indecision			
	Decreased Appetite	.50		-.38
	Weight loss	.42		
	Increased Appetite			
	Weight gain			
	Psychomotor agitation			
	Psychomotor retardation			
	Worthlessness	.47		
	Guilt			
	Hopelessness	.49		
Bipolar Disorder	Elevated mood			
	Decreased need for sleep			
	Increased goal-directed activity			
	Racing thoughts	.35		

K-SADS-PL		MMPI-A-RF H-O Scales		
Section	K-SADS-PL item stem	EID	THD	BXD
	Grandiosity			
	Pressured speech			
	Poor judgment			.51
	Distractibility			
	Physical restlessness			
	Influence of alcohol or drugs	-.35		.41
Psychosis	Hallucinations		**.53**	
	Delusions		.43	
	Flat affect		.41	
	Inappropriate affect		.50	
	Incoherence		.52	
	Loosening of associations		.52	
	Catatonic behavior		.52	
Panic Disorder	Panic attacks	.50		
	Fear of another attacks	.49		
	Agoraphobia	.53		
	Maladapt. changes to avoid attacks			
GAD	Excessive worry	.51		
	Overconcern with competence	.36		
	Need for reassurance			
	Ability to control worries	.43		
	Muscle tension	.44		
	Restlessness	.45		
	Easily fatigued	.47		
	Feeling keyed up	.42		
	Difficulty concentrating	.43		
	Trouble falling or staying asleep			
	Irritability			
PTSD	Recurrent thoughts or images			
	Efforts to avoid thoughts or feelings			
	Nightmares			.36
	Insomnia			.39

	K-SADS-PL	MMPI-A-RF H-O Scales		
Section	K-SADS-PL item stem	EID	THD	BXD
	Irritability			
	Re-enactment			
	Dissociative episodes			
	Distress to exposure			
	Efforts to avoid physical reminders			
	Inability to recall details			
	Diminished interest in activities			
	Detachment			.37
	Restricted affect			
	Sense of foreshortened future			
	Difficulty concentrating			
	Hypervigilance			
	Exaggerated startle response			
	Physiological reactivity			
	Flashbacks			
	Negative beliefs			
	Distorted cognitions			
	Negative emotional state			
	Inability to experience positive emotions			
	Reckless or self-destructive behavior			.43

Note: Large correlations (≥ .53) are bolded. EID = Emotional/Internalizing Dysfunction; THD = Thought Dysfunction; BXD = Behavioral/Externalizing Dysfunction; MDD = Major Depressive Disorder; GAD = Generalized Anxiety Disorder; PTSD = Posttraumatic Stress Disorder.

Table 19

Substantial K-SADS-PL Externalizing Item-Level Correlates (≥ .35) for the MMPI-A-RF H-O Scales

	K-SADS-PL	MMPI-A-RF H-O Scales		
Section	K-SADS-PL item stem	EID	THD	BXD
	Difficulty sustaining attention			
ADHD	Easily distracted			
	Difficulty remaining seated			

K-SADS-PL		MMPI-A-RF H-O Scales		
Section	K-SADS-PL item stem	EID	THD	BXD
	Impulsivity			.38
	Careless mistakes			
	Doesn't listen			
	Difficulty following instructions			
	Difficulty organizing tasks			
	Dislikes tasks requiring attention			
	Loses things			
	Forgetful			
	Fidget			
	Runs or climbs			
	On the go			
	Difficulty playing quietly			
	Blurts out answers			
	Difficulty waiting turn			
	Interrupts or intrudes			
	Shifts activities			
	Talks excessively			
	Engages in dangerous activities			.43
ODD	Loses temper		.38	.51
	Argues a lot with adults			.39
	Disobeys rules			**.61**
	Easily annoyed			.37
	Angry or resentful			
	Spiteful or vindictive		.35	.46
	Uses bad language	-.44		.46
	Annoys people on purpose			.45
	Blames others for mistakes			
Conduct Disorder	Lies			.45
	Truant			.44
	Initiates physical fights	-.42		**.60**
	Bullies, threatens, or intimidates		.42	.45
	Nonaggressive stealing			**.57**
	Vandalism			**.59**
	Breaking and entering	-.35		.52
	Aggressive stealing			**.58**
	Firesetting			

K-SADS-PL		MMPI-A-RF H-O Scales		
Section	K-SADS-PL item stem	EID	THD	BXD
	Often stays out at night			**.62**
	Ran away overnight			**.67**
	Use of a weapon			**.54**
	Physical cruelty to others			.51
	Forced sexual activity			
	Cruelty to animals			
Alcohol Use	Frequency			.52
	Concern from others			**.56**
	Drinks more than planned			**.53**
	Negative physical consequences			.52
	Dangerous behavior			**.63**
	Negative psych. consequences			**.58**
	Negative occupational consequences			**.62**
	Negative social consequences			**.64**
	Legal consequences			**.54**
	Failure to fulfill major obligations			**.57**
	Important activities given up			**.71**
	Time consuming			**.55**
	Tolerance			**.57**
	Tried to quit			**.58**
	Withdrawal			**.57**
	Alcohol consumed to relieve withdrawal			**.60**
Substance Use	Frequency			**.59**
	Uses more than planned			**.53**
	Negative physical consequences			**.61**
	Dangerous behavior			**.67**
	Negative psych. consequences			**.56**
	Negative occupational consequences			**.61**
	Negative social consequences			**.56**
	Legal consequences	-.35		**.66**
	Failure to fulfill major obligations			**.56**
	Important activities given up			**.61**
	Time consuming			**.61**
	Tolerance			**.57**
	Tried to quit			.50
	Withdrawal			**.56**

K-SADS-PL		MMPI-A-RF H-O Scales		
Section	K-SADS-PL item stem	EID	THD	BXD
	Drugs consumed to relieve withdrawal			**.58**

Note: Large correlations (≥ .53) are bolded. EID = Emotional/Internalizing Dysfunction; THD = Thought Dysfunction; BXD = Behavioral/Externalizing Dysfunction; ADHD = Attention Deficit/Hyperactivity Disorder; ODD = Oppositional Defiant Disorder.

Supplementary Analyses

Supplementary Research Question 1: Are MAYSI-2 subscales effective at screening DSM-5 symptomatology?

Given the MAYSI-2's frequent use in juvenile justice settings (Grisso, 1999), it is crucial to empirically test its effectiveness at identifying potential psychological difficulties. The MAYSI-2 also provides warning and caution classifications intended to alert mental health professionals to clinical-diagnostic issues that may need professional intervention. Thus, this supplementary research question aimed to formally test the efficacy of the caution classifications of the MAYSI-2. Specifically, a series of one-way ANOVAs compared the corresponding K-SADS-PL diagnostic section total scores between adolescents falling into the three MAYSI-2 classification categories.

Broadly, adolescents meeting warning threshold tended to produce significantly more symptomatology than those without. This trend held true across nearly all MAYSI-2 scales and relevant K-SADS-PL diagnostic sections. For example, the Alcohol/Drug Use MAYSI-2 scale produced particularly large differences (*M d* = 2.17) between juveniles falling in the warning category and those not meeting threshold for any classification.

Table 20

Differences in Relevant K-SADS-PL Symptomatology Total Scores by Caution/Warning Range Classifications for MAYSI-2 Scales

MAYSI-2 Scale	K-SADS-PL	MAYSI-2 Classification						*F*	*p*	*d1*	*d2*	*d3*
		None		Caution		Warning						
		M	*SD*	*M*	*SD*	*M*	*SD*					
Alcohol/	Alcohol	8.98_{a}	13.47	12.67_{ab}	19.40	31.86_{b}	15.61	8.09	**.01**	1.56	0.22	1.10
Drug Use	Drug	7.20_{a}	12.90	27.67_{b}	6.51	32.57_{b}	3.41	16.44	**<.001**	2.78	1.96	0.96
Angry-Irritable	ODD	12.74_{a}	4.78	12.96_{a}	5.63	20.75_{b}	3.37	8.45	**.01**	1.96	0.04	1.71
	CD	19.96_{a}	5.30	17.24_{b}	6.77	25.00_{a}	6.78	4.87	**.01**	0.82	0.45	1.15
Depressed-	MDD	42.92_{a}	8.27	54.06_{b}	8.74	62.62_{c}	7.25	18.35	**<.001**	2.55	1.31	1.07
Anxious	Panic	2.08_{a}	0.64	2.63_{a}	1.35	4.00_{b}	1.00	9.98	**<.001**	2.25	0.53	1.17
	GAD	15.92_{a}	3.71	17.61_{a}	2.70	22.07_{b}	2.69	15.66	**<.001**	1.92	0.52	1.66
Suicide Ideation	MDD	40.35_{a}	5.00	60.00_{b}	1.41	59.00_{b}	10.60	53.24	**<.001**	2.20	5.19	-0.13
Thought Disturbance[a]	Psychosis	2.00_{a}	0.00	23.00_{a}	29.69	26.33_{b}	20.23	7.04	**.01**	1.64	1.04	0.13

Note. Means sharing a common subscript are not statistically different at $p < .05$ according to the Tukey HSD procedure. *d1* compares None with Warning; *d2* compares None and Caution; *d3* compares Caution and Warning. MDD = Major Depressive Disorder; GAD = Generalized Anxiety Disorder; ODD = Oppositional Defiant Disorder; CD = Conduct Disorder; Alcohol = Alcohol Use Disorder; Drug = Substance Use Disorder. [a] Per the MAYSI-2 manual, the Thought Disturbance scale is specific to boys only.

Unexpectedly, the relationship among no classifications, caution classifications, and warning classifications was not always linear. Conceptually, it stands to reason that those without any classification would exhibit the lowest symptomatology, those in the caution classification significantly more symptomatology, and those in the warning classification the most. Yet, in several instances, the caution classification category was not significantly different from either of the other classification groups. To illustrate, on the Depressive-Anxious MAYSI-2 subscale, the caution classification produced only minimally more anxious symptomatology than the no classification category. In another instance, the warning classification on the Angry-Irritable MAYSI-2 subscale did not exhibit significantly more CD symptoms than those without any classification.

Suicidal ideation among juveniles is obviously a very serious concern; therefore adequate identification of such thinking is critical. As seen in Table 21, MAYSI-2 Suicide Ideation scale warning and caution classifications identified youth with significantly elevated suicidal ideation. Encouragingly, the caution and warning classifications appeared to work well. That is, both classifications exhibited a perfect specificity, while maintaining excellent sensitivity. Consequently, clinicians can feel comfortable ruling out true suicidal ideation—at least empirically—when the suicidal ideation scale is below any classification threshold.

Table 21

Effectiveness of MAYSI-2 Caution/Warning Classifications for Suicidal Ideation Scale

Cut Score	Sensitivity	Specificity	OCC	Base Rate [a]	
				PPP	NPP
Caution ≥ 2	.87	1.00	.89	1.00	.98
Warning ≥ 3	.83	1.00	.86	1.00	.98

Note: OCC = overall correct classification; PPP = positive predictive power; NPP = negative predictive power. [a] Following recent adolescent epidemiological research (Nock et al., 2013), a 12.1% base rate was used.

Supplementary Research Question 2: Does the absence of clinical elevations on MMPI-A-RF RC scales (i.e., WNL profiles) provide any clinical correlates?

As previously discussed in the Introduction, one common criticism of the original MMPI-A was the relatively high frequency of profiles without any substantial elevations, which are called "within normal limit" (WNL) profiles. While it has been argued that the design of the MMPI-A-RF may help to reduce WNL profiles (Stokes, Pogge, & Archer, 2018), it is important to evaluate this assertion by examining the prevalence of WNL profiles among MMPI-A-RF protocols. Thus, as an initial step, the prevalence of WNL profiles in the current study was evaluated. Of 66 valid profiles, 9 (13.6%) did not evidence any significant elevations on any of the MMPI-A-RF RC scales. Given the small sample size, it is imperative to view these initial analyses of WNL profiles as preliminary.

Next, analyses helped to determine whether adolescents with WNL profiles truly lacked significant mental health issues or whether the MMPI-A-RF was unable to identify them. More specifically, a series of one-way ANOVAs compared the number of K-SADS-PL symptoms meeting clinical threshold between adolescents with and without WNL profiles. Encouragingly, all four statistically significant differences were in the correct direction, with those without WNL profiles exhibiting considerably more MDD (d = 1.16), panic (d = 0.71), GAD (d = 1.03), and PTSD (d = 3.67) symptoms. Unsurprisingly, no significant differences between subsamples existed in terms of manic or psychotic symptomatology due to the low base rates of these symptoms in the current sample.

Against expectations, adolescents with and without WNL profiles exhibited similar numbers of clinically significant externalizing symptoms. As a startling statistic, multiple adolescents met clinical threshold on over half of the ADHD (i.e., 4 people endorsed $\geq$ 9 out of 17 total ADHD symptoms) or ODD symptoms (i.e., 3 people endorsed $\geq$ 5 out of 9 total ODD

symptoms). Arguably, ADHD symptomatology is not thoroughly covered by MMPI-A-RF items, and thus would have a difficult time identifying adolescents who are exhibiting clinically threshold symptoms. On the other hand, previous analyses have demonstrated the robust properties of the RC4 scale in identifying ODD symptoms. Thus, it stands to reason that adolescents reporting a significant number of ODD symptoms should produce an elevation on RC4.

Table 22

Preliminary Analysis of Substantial Correlations (≥ .35) of MMPI-A-RF RC scales with the K-SADS-PL Total Scores for WNL Profiles (n = 9)

	K-SADS-PL Total Scores										
	Internalizing						Externalizing				
	MDD	Mania	SCZ	Panic	GAD	PTSD	ODD	CD	ALC	DRG	ADHD
RCd	**.61**	.40			**.73**		.35			-.35	**.55**
RC1			**-.65**		.36			-.47	-.35		
RC2	**.81**		.35		**.81**	-.36		-.39	-.41	**-.56**	.52
RC3	**-.63**		-.44			**.59**			**.69**		-.39
RC4		.39	-.63		-.38	**.66**			**.56**	**.87**	
RC6	**.57**										
RC7	**.69**	.49			.50		**.58**				**.80**
RC8	**.73**	**.80**		.40			.45		.38		.37
RC9	.45	.36			**.62**						

Note: Large correlations (≥ .53) are bolded. RCd = Demoralization; RC1 = Somatic Complaints; RC2 = Low Positive Emotions; RC3 = Cynicism; RC4 = Antisocial Behavior; RC6 = Ideas of Persecution; RC7 = Dysfunctional Negative Emotions; RC8 = Aberrant Experiences; RC9 = Hypomanic Activation; MDD = Major Depressive Disorder; SCZ = Schizophrenia/Psychosis; GAD = Generalized Anxiety Disorder; PTSD = Posttraumatic Stress Disorder; ODD = Oppositional Defiant Disorder; CD = Conduct Disorder; ALC = Alcohol Use Disorder; DRG = Substance Use Disorder; ADHD = Attention Deficit/Hyperactivity Disorder.

Focusing on the interpretability of the WNL profiles, their clinical correlates were explored. As seen in Table 22, WNL profiles produced a more diffuse pattern of clinical correlates compared to the entire study sample. That is, MMPI-A-RF RC scales frequently

exhibited at least moderate correlations with multiple K-SADS-PL diagnostic sections. In several instances, several unexpected clinical correlates were observed. For instance, in WNL profiles, RC8 correlates largely with mood sections (e.g., MDD = .73), moderately with some externalizing sections (e.g., ADHD = .37), but insignificantly with psychotic symptomatology. Furthermore, several MMPI-A-RF RC scales appeared substantially associated with ADHD symptoms, despite previous analyses with the full study sample failing to find any such relationships with ADHD symptomatology. Given the very low sample size of WNL profiles in the current data, no concrete interpretations can be made. Additional research will help to confirm whether this relationship was obscured in the full study sample or if it is merely a spurious relationship.

Supplementary Research Question 3: Do elevations on the MMPI-A-RF Antisocial Behavior scale predict aggressive conduct problems in youth with mental health needs?

The third supplementary research question sought to investigate the utility of the MMPI-A-RF in predicting aggressive conduct problems in youth with mental health needs.

Table 23

MMPI-A-RF RC4 and AGG-r Scales as Predictors of Aggressive Conduct Problems

	Beta	SE	β	*t*	*p*	Partial *r*
RC4	.10	.02	.51	4.29	**<.001**	.48
AGG-r	.03	.02	.16	1.37	.18	.17

Note. RC4 = Antisocial Behavior; AGG-r = Aggressiveness-Revised.

A multiple regression determined the relative efficacy of the MMPI-A-RF Antisocial Behavior scale and Aggressiveness-Revised scale in predicting aggressive conduct problems as measured by the K-SADS-PL. Taken together, the MMPI-A-RF variables accounted for 36.5%

of the variance. Of the two variables, RC4 emerged as the only significant predictor. Furthermore, the standardized beta weights demonstrated that RC4 contributes the most to the prediction of conduct problems (see Table 23).

CHAPTER 4

DISCUSSION

Clinical correlates constitute an essential, although sometimes overlooked, part of the interpretation process for multiscale inventories. Importantly, such correlates empirically provide additional meaning and context to test scores, above and beyond scale descriptions. The following section provides a brief overview of how clinical correlates of the MMPI family have been interpreted historically.

In a seminal investigation of MMPI codetypes, Gynther, Altman, and Sletten (1973) argued there were two fundamental approaches to interpretation of MMPI correlates: actuarial and non-actuarial. According to Gynther et al. (1973), the major difference between actuarial and non-actuarial approaches to interpretation relies on the source of the correlates, with actuarial approaches relying on empiricism and non-actuarial approaches relying instead on clinical experience. Moreover, the authors openly questioned the validity of a non-actuarial approach and instead opted to develop a comprehensive actuarial system for the interpretation of MMPI profiles. Using the Missouri Standard System of Psychiatry (SSOP; Sletten, Ulett, Altman, & Sundland, 1970), Gynther et al. systematically examined the relationship between SSOP items and MMPI codetypes. Such methodology established the general paradigm for ascertaining empirical clinical correlates going forward (Rogers et al., in press).

Following the initial establishment of MMPI clinical correlates, empirically based clinical correlates have been established for all subsequent iterations of the MMPI. In his work on the interpretation of the MMPI-2/MMPI-2-RF, Greene (2011) published an amalgamation of research on the MMPI, MMPI-2, and MMPI-2-RF over the last 50 years. Included in this reference book were specific clinical correlates for each MMPI-2 and MMPI-2-RF scale, as well

as MMPI-2 codetypes. In the following year, Ben-Porath (2012) elected to include specific empirical correlates for each MMPI-2-RF substantive scale in his most recently published guide to interpretation of the MMPI-2-RF. Published MMPI-2/MMPI-2-RF empirical correlates stem from multiple research studies (e.g., Graham, Ben-Porath, & McNulty 1999; Tellegen & Ben-Porath, 2008) ranging in the rigor of their external criteria. For example, several studies (e.g., Graham et al., 1999) simply elected to utilize self-report inventories, such as the SCL-90-R (Derogatis, 1983), as the external criterion. The similarity between such self-report inventories and the MMPI-2 potentially increased the likelihood of finding similar correlates (Greene, 2011). It could also be argued that asking the same source similar questions does not truly constitute independent, external criteria. Alternatively, other researchers (e.g., Archer et al., 1995) relied on clinical rating forms completed by collateral sources, such as a BPRS completed by treating psychologists. On this point, Greene (2011) convincingly argued this methodology provides more independent estimates of correlates.

In the adolescent literature, MMPI-A clinical correlate studies, and by extension MMPI-A-RF studies, vary considerably in the rigor of external criteria. As discussed in the Introduction, the three primary studies cited in the MMPI-A-RF manual (i.e., Forbey & Ben-Porath, 2003; Handel et al., 2011; Veltri et al., 2009) utilized three very different external assessment measures for developing clinical correlates. Interestingly, one common critique observed in all three studies was the minimal threshold used for determining clinical correlates. Handel et al. (2011) opted to utilize .20 as the threshold for "substantively meaningful" (p. 572) correlates. In addition to avoiding the inclusion of zero in 95% confidence intervals, the authors argued such a small magnitude was in line with previously conducted correlational research on the MMPI-A. Nevertheless, justifying the use of a relatively lax threshold simply in order to be consistent with

prior research appears to be faulty reasoning. Veltri et al. (2009) took a different approach and looked to alpha to inform their threshold for clinical correlate significance. A Bonferroni correction determined significance to be $p < .002$. Coincidentally, such an alpha translated to correlational thresholds largely consistent with Handel: .20 in the forensic sample and .22 in the inpatient sample.

While an important first step, previous investigations of the empirical clinical correlates of the MMPI-A-RF have several notable methodological weaknesses. First, previous studies have generally utilized external assessment measures of either insufficient depth (i.e., behavior checklists; Handel et al., 2011) or a novel but unvalidated measure (Forbey & Ben-Porath, 2003). Second, all previous studies have used relatively lax criteria for determining clinical correlates (e.g., correlations $\geq .20$). The current study aimed to rectify this literature by examining the clinical correlates of the MMPI-A-RF employing a well-validated, structured interview (i.e., K-SADS-PL) that assesses a wide range of psychopathology and applying rigorous thresholds for clinical significance.

This chapter is organized into four major sections that serve to contextualize the current study's findings. The first section provides an overview of the psychometric properties of the MMPI-A-RF as an emerging psychological adolescent assessment measure. Specifically, reliability and validity and clinical correlates are discussed. The second section critically examines gender, including how it may affect interpretation of the MMPI-A-RF. The third section integrates the previous two sections by focusing on practical clinical applications. Finally, the fourth section addresses limitations of the current study and outlines future directions for continued research.

The Clinical Utility of the MMPI-A-RF

The MMPI-A-RF was originally developed with the aim of improving upon the previously published, and widely used, MMPI-A. Specifically, Archer and his colleagues (2016) sought to improve upon the psychometric properties over its predecessor, the MMPI-A, in terms of content heterogeneity and discriminant validity. The MMPI-A-RF manual provided the initial results regarding whether this aim was met. However, continued validation of the MMPI-A-RF's psychometric properties is imperative.

Scale Homogeneity of the MMPI-A-RF Reconstructed Clinical Scales

Investigating scale homogeneity and reliability of its RC scales was the first systematic step in investigating the MMPI-A-RF's psychometric properties. In the current study, the internal consistencies of the RC scales were rather variable, ranging from poor to excellent. While findings from the current study generally mirrored the previously documented internal consistencies for RC scales (see Archer et al., 2016), notable differences emerged. Of note, internal consistencies were reexamined separated by gender in order to make direct comparisons with those listed in Archer et al. (2016); see Table A.4 in the Appendix.

Interestingly, a pattern emerged in which longer RC scales (i.e., consisting of 10 or more items) manifested generally good scale homogeneity across the literature (e.g., Forbey & Ben-Porath, 2003; Handel et al., 2011; Veltri et al., 2009). In contrast, shorter scales (9 items or less) evidenced either poor or highly variable psychometric properties. As a major strength, RCd, RC1, RC2, and RC4 all demonstrated highly convergent (i.e., < .10 differences in alpha) internal consistencies across the MMPI-A-RF literature (Archer et al., 2016) with the findings of the

current study included. Moreover, three of those scales exhibited internal consistencies that met the Clark and Watson (1995) threshold for strong homogeneity.

In stark contrast, RC9 has consistently evidenced the poorest internal consistency among RC scales. Poor alphas (M α = .56) have previously been noted in outpatient, inpatient, correctional, and school settings (see Archer et al., 2016). Alpha for RC9 in the current study was even lower (α = .47), yet still generally consistent with previous research. Such a poor alpha is indicative of over 50% non-systematic error (Henson, 2001). Contributing to the low alpha, the current study also found consistently poor inter-item correlations (M r = .10) for RC9. Taken together, such poor reliability does not meet the established threshold for adequate reliability (Clark & Watson, 1995; Miles, Fulbrook, & Mainwaring-Magi, 2018). Consequently, RC9 should be considered for scale refinement. Toward this goal, the current study sought to identify and remove items correlating poorly with the overall scale. Unfortunately, only a minimal improvement in alpha (i.e., .07) was achieved with the removal of a poorly related item, at the cost of shortening an already brief scale.

Alphas on the remaining RC scales (i.e., RC3, RC6, RC7, and RC8) were generally lower in the current study than those previously documented in the extant literature (Archer et al., 2016). When compared to a strictly outpatient sample (Archer et al., 2016)—the sample most analogous to the current study sample—notable differences still remained. While slightly different in males, the more prominent differences in alphas occurred in females. As the most extreme example, alphas for RC7 differed by nearly .20 in females, decreasing from .75 (i.e., Archer et al., 2016) to .56 (i.e., current study). One hypothesis for the gender difference was related to a restricted range for males. In inspecting the standard deviations and ranges, however, males and females were comparable.

Construct Validity of the MMPI-A-RF Reconstructed Clinical Scales

Correlational analyses from the current study and previous research (e.g., Handel et al., 2011) provided some generally consistent evidence of convergent and discriminant validity of the MMPI-A-RF. According to the MMPI-A-RF manual (Archer et al., 2016), evidence of the instrument's construct validity has been demonstrated with a wide range of extra-test indicators, on both scale and item levels. For the purposes of these analyses, only scale level correlates were analyzed (see subsequent section on interpretive statements for an investigation of item level correlates). As such, Handel et al. (2011) was the only previous study directly comparable to the current investigation.

Unlike the current study, Handel et al. (2011) utilized a battery of symptom and behavior checklists as their external criteria. As previously described, adolescents completed the Youth Self-Report (YSR; Achenbach & Rescorla, 2001), while collateral sources completed the Child Behavior Checklist (CBCL; Achenbach & Edelbrock, 1983) and the Disruptive Behavior Rating Scale (DBRS; Barkley & Murphy, 1998). Across all Handel et al. (2011) study measures, scale scores were computed to reflect broad domains of psychopathology, such as somatic complaints, thought disturbance, and acting out behaviors. Such general domains arguably paralleled the K-SADS-PL diagnostic section total scores used in the current study. Moreover, the scale scores of the YSR were most relevant to the current study given that the adolescent reported on their own psychopathology directly via these measures, instead of using a collateral source. As previously addressed, Handel et al. (2011) utilized a lowered threshold of .20 as significant; however, the following discussion focused on relationships that met or exceeded the more stringent threshold for the current study (i.e., $\geq$ .35).

Across both studies, RC2 did not evidence the expected relationship with conceptually related depressive symptomatology. In fact, RC2 evidenced only moderate correlations with depressive symptomatology (r = .36 in the current study and .43 in Handel et al., 2011) among females. Given the focus of RC2, this was particularly surprising. Taken together, a reexamination of the content of RC2 scale items may be warranted to better reflect the psychopathology conceptually related to its content. For example, current RC2 items center around feelings of pessimism, anhedonia, and social isolation. It is possible that creating items reflecting classic DSM-5 symptoms of depression, such as depressed mood, feelings of worthlessness, and thoughts about death, may improve RC2's convergent and discriminant validity.

Poor convergent validity markedly limits RC3's clinical utility. RC3 failed to produce any substantial correlations ($\geq$ .35) with either K-SADS-PL diagnostic section totals or YSR scales, reflecting poor convergent validity in both studies. Consequently, additional examination into the psychometric properties and content of RC3 is warranted. Further examination of RC3's items revealed two items (i.e., item 31 and 131) yielding poor inter-item correlations with fellow scale items and poor correlations with the overall scale. However, if removed, alpha increased negligibly (i.e., merely .01 in either case). Given the already low number of items on RC3, it does appear prudent to remove these items. Instead, one option to improve RC3's psychometric properties would be to replace these poorly functioning items.

In the remaining RC scales, Handel et al. (2011) evidenced more diffuse patterns of convergent validity and diminished patterns of discriminant validity than the current study findings. In terms of convergent validity, certain RC scales in Handel et al. (2011) correlated at least moderately with the majority of YSR scales. For example, RCd (M = .55) and RC7 (M =

.46) evidenced at least moderate correlations with nearly all internalizing scales of the YSR. Yet, RCd and RC7 also evidenced moderate correlations with several externalizing scales of the YSR, an unexpected finding that limited their discriminant validity (Handel et al., 2011). By comparison, the current RCd and RC7 both only correlated with MDD, GAD, and panic symptomatology in the internalizing domain.

Contrary to the current findings, RC9 generated some evidence of convergent and discriminant validity in Handel et al.'s (2011) study. They found RC9 was correlated with affective symptoms (M = .36), thought problems (M = .39), and internalizing symptoms in general (r = .38 in males) in the internalizing domain. Additionally, RC9 was correlated with ADHD symptoms (M = .38 in females), aggressive behaviors (M = .39), and externalizing symptoms in general (r = .41 in males). Clearly, further research is needed in light of these highly discrepant findings.

On a positive note, current findings provide stronger and cleaner patterns of convergent and discriminant validity for RC4 than those documented by Handel et al. (2011). First, in the current study, RC4 evidenced large to very large correlations for ODD, CD, and substance abuse sections of the K-SADS-PL. To the contrary, Handel et al. (2011) observed only moderate to large relationships with regard to DSM ODD and CD scales of the YSR. As an interesting exception, Handel et al.'s (2011) RC4 also evidenced moderate correlations with inattention symptoms (M = .39). Unexpectedly, this finding was not replicated in the current study, despite an entire K-SADS-PL section dedicated to ADHD symptoms. Second, the only internalizing K-SADS-PL diagnostic sections to be substantially related to RC4 were mania and PTSD, two internalizing domains in which certain acting out behaviors are conceptually related. In contrast, Handel et al.'s (2011) RC4 exhibited more limited discriminant validity with unexpected

moderate correlations with several YSR scales related to affective, anxiety, thought problems, and internalizing problems in general. One initial hypothesis for explaining these relationships was that such YSR scales included items related to the acting out behaviors of mania and PTSD that had been documented in the current study. Yet, a further examination of YSR scale content disproved this hypothesis.

Most strikingly, Handel et al. (2011) found virtually no convergent validity between the RC scales and either the CBCL or DBRS, when held to the present rigorous standard ($r \geq .35$). Surprisingly, only two scales evidenced at least moderate relationships with RC scales. For instance, the Rule-Breaking Behavior scale was the only externalizing scale to evidence moderate correlations with RC4. This finding stands in stark contrast to the myriad of relationships found among K-SADS-PL diagnostic totals of the current study and YSR scales in the same study (Handel et al., 2011). Such differences may reflect the source of reporting. As reviewed in the Introduction, De Los Reyes and Kazdin (2005) documented discrepancies between sources of information. According to previous research (Cantwell et al., 1997), adolescents tended to report more internalizing symptoms, while collateral sources generally reported more observable externalizing symptoms. Yet, in Handel et al. (2011), adolescents reported more symptomatology than collateral sources across the board. Thus, it is possible the collateral sources did not know the extent of the psychopathology of many adolescents.

Examination of Interpretive Statements for Reconstructed Clinical Scales

Archer et al. (2016) included in the MMPI-A-RF manual interpretive statements to augment interpretation of elevated RC scale scores. According to those authors, the interpretive statements stem from previous research (e.g., Handel et al., 2011; Veltri et al., 2009) on the

MMPI-A-RF's empirical correlates. However, empirical correlates are more encompassing than those listed. Instead, the MMPI-A-RF authors included the more comprehensive list of item-level correlational data from MMPI-A-RF studies in their appendix for review. With a goal of examining empirical correlates of the RC scales, item-level analyses were also conducted in the current investigation.

In terms of empirical correlates, RC4 was arguably the most well-replicated scale across MMPI-A-RF studies. The current study identified numerous correlates across a range of externalizing disorders (e.g., ODD and CD), demonstrating excellent convergent validity. As demonstrated in Table A.5 in the current Appendix, a variety of such correlates replicated the findings of previously conducted studies (Archer et al., 2016; Handel et al., 2011). As one cogent example, empirical correlates related to involvement with criminal behavior were identified across all studies. Yet, the level of specificity of the correlates ranged dramatically. For example, Veltri and colleagues (2009) only vaguely identified a history of criminal charges as being a moderate clinical correlate with RC4. Meanwhile, other research, including the current study, documented a wide range of correlations (i.e., virtually zero to large) among specific criminal charges, indicating that certain criminal behavior may be driving the relationship. While certain researchers (Forbey & Ben-Porath, 2003) agreed with the current study findings that fire setting, forced sexual activity, or cruelty to animals were virtually unrelated ($rs \leq .35$) to RC4, there was disagreement among other charges. Such a disparity in relationships may stem from differences in legal activity among the study samples, with Forbey and Ben-Porath's (2003) sample primarily engaging in status offenses and the current sample primarily engaging in property offenses.

In the same vein, current results expanded upon previous findings by documenting RC4's relationship across other disorders. Again, a previous study (Forbey & Ben-Porath, 2003) ambiguously recognized a history of substance abuse as an empirical correlate of RC4 (r = .35). Yet, the current correlations range from large to very large for a diverse range of behaviors across both alcohol and substance abuse, with some of the largest correlations relating to particularly negative consequences of serious abuse (i.e., r = .74 for giving up important activities to consume alcohol). Thus, particularly high scores on RC4 may be indicative of not just substance abuse but particularly severe substance abuse that may require clinical interventions.

Despite general convergence with previous literature in terms of RC4 correlates, some empirical correlates in the current study had not been previously recognized. Specifically, several PTSD features were at least moderately correlated to RC4 in the current study. Importantly, the use of non-PTSD specific external criterion measures in previous studies mitigates most direct comparisons of PTSD features with RC4; yet, some criteria can be indirectly compared. For example, Veltri et al. (2009) found largely weak convergence (rs between .00 and .23) between RC4 and PTSD-like symptoms, namely irritability, sleeplessness, and nightmares. This finding is in direct contrast to the current findings of moderate correlations for all three of those symptoms. Researchers might assume that the lack of convergence between such studies may be the differences in trauma exposure among the samples. While the majority of the current sample reported at least some exposure to a traumatic event, it is not clear what proportion of Veltri et al.'s (2009) sample was similarly exposed. Thus, it is possible that Veltri et al.'s (2009) sample was not experiencing clinical levels of trauma symptoms. Such a restricted range of trauma endorsement may explain why such correlations did not exist among PTSD symptoms and RC4.

The MMPI-A-RF clinical correlate literature (Archer et al., 2016) has also found good convergence among RC8 and psychotic symptomatology. Current study results mirrored this pattern of excellent convergent and discriminant validity for RC8 by identifying empirical correlates exclusively in the psychosis domain. Across multiple studies (Forbey & Ben-Porath, 2003; Veltri et al., 2009), strong relationships have been documented between hallucinations and elevated RC8 scores. As one unique contribution to the MMPI-A-RF literature, the current study was able to extend the empirical correlates of RC8 beyond positive psychotic symptoms by identifying moderate to large correlations (M = .54) with negative psychotic symptomatology; thus, bolstering support for RC8's convergent validity.

Strong patterns of convergent validity were observed for RCd in the current investigation. The current findings found moderate to very large correlations for MDD, GAD, and panic symptomatology. Such correlates were consistent with Archer and colleagues' (2016) RCd empirical correlates which clustered around depressive (e.g., suicidal ideation and low self-esteem) and anxious features. As a unique finding, the current study found additional evidence for RCd's convergent validity by identifying moderate, bordering on large, correlations with worthlessness (r = .45) and hopelessness (r = .49).

In terms of RCd's discriminant validity, previous research (Archer et al., 2016; Veltri et al., 2009) had documented contrasting findings related to problems with attention and concentration. In particular, Veltri et al. (2009) reported a moderate relationship (r = .43 in males) with acute psychiatric inpatients, but failed to find a commensurate relationship in a forensic setting (r = .11). The current findings were consistent with the latter relationship. Scholars may be concerned that inattentive symptoms were not adequately measured, but the K-SADS-PL assesses both inattentive and hyperactive ADHD symptoms. Another hypothesis

might be that ADHD symptoms were not prominent in the current sample, but this seems unlikely because nearly one quarter of the sample (n = 20) received an ADHD diagnosis.

Unlike the previously described RC scales, RC3 had only one clinical correlate in the current study. It addressed consuming alcohol to relieve withdrawal symptoms. Likewise, previous investigations have generally failed to identify even a few empirical correlates. For example, Archer et al. (2016) found only the engagement in rule-breaking behavior. Yet, in the current study, no other items related to rule-breaking behavior evidenced substantial correlations with RC3. Similarly, other past studies examining RC3 have failed to find substantial correlations ($\geq$ .35; Forbey & Ben, Porath, 2003; Veltri et al., 2009). Such poor convergent and discriminant validity across the literature constrains RC3 interpretation at least from an empirical perspective.

Like RC3, RC9 also has limited empirical correlates. Encouragingly, Archer and colleagues (2016) identified the following RC correlates: aggressive behavior, a history of conduct problems, and having numerous interests. Yet, other empirical investigations, including the current study, were not able to replicate these findings. In fact, the current study found RC9 had only one empirical correlate with the ADHD feature of "constantly being on the go" (r = .35). Unfortunately, other studies were not successful at the .35 threshold for a substantial relationship. Importantly, the lack of scale homogeneity for RC9 may directly contribute to the very limited number of empirical correlates. In the absence of replicated empirical correlates, a more traditional, non-empirical approach to interpretation is recommended.

Other RC scales' clinical correlates found more mixed findings in the context of the overall MMPI-A-RF clinical correlate literature. Consistent with its content, RC1 was intended to assess for a wide range of somatic complaints. Of its two empirical correlates, Archer et al.

(2016) found one predicted (i.e., multiple somatic complaints) and one unexpected (i.e., problems with concentration and attention) associations. As previously noted, the current study did not assess somatic complaints directly. Instead, general convergent validity was found when examining correlates along the internalizing spectrum. An examination of correlates with a somatic focus yielded several positive findings: sleep disturbance (e.g., non-restorative sleep, r = .39), fatigue (r = .42), and the effects of excessive worry (e.g., muscle tension, r = .47). More generally, other empirical correlates involved depressive and anxious symptomatology within the internalizing domain.

Due to its focus on low positive emotions, RC2 would be expected to evidence convergent validity through many empirical clinical correlates related to depressive symptomatology. Archer et al.'s (2016) list of empirical correlates for RC2, such as anhedonia, hopelessness, depressed mood, and being self-punishing, were consistent with convergent validity. As a very positive result, correlates of anhedonia and hopelessness were replicated in this dissertation. Unfortunately, other key symptoms of depression, such as suicidal ideation and depressed mood, were not. Prior research (Veltri et al., 2009) also generally failed to find at least moderate ($\geq$ .35) correlations for typical symptoms of depression, limiting RC2's convergent validity. As a gender-specific exception, Forbey and Ben-Porath (2003) established the following RC2 correlates for females in residential treatment: depression, low self-esteem, and suicide attempts or gestures. As previously noted, RCd demonstrated generally stronger relationships with depressive symptomatology than RC2. Thus, it may be helpful for clinicians to also take into account elevations on RCd when evaluating whether an adolescent is struggling with depressive symptoms.

In the current study, RC6 yielded several moderate correlations with psychotic symptomatology and oppositional behavior, strengthening its convergent validity. According to Archer et al. (2016), elevated RC6 scores were empirically correlated to psychotic symptoms (e.g., auditory or visual hallucinations) or acting out behaviors (e.g., aggressive or oppositional behaviors). In the present study, RC6 correlates were expanded beyond hallucinations to also include salient negative psychotic symptoms. Also, like Archer et al. (2016), several moderate correlations were found between RC6 and oppositional behavior, such as losing temper, being spiteful or vindictive, and bullying others. While not a perfect match, these two studies still provide strong convergent validity. Nevertheless, it is important for psychologists to look at RC4 and/or RC8 to better understand the nature of an adolescent's elevated RC6 score. Otherwise, psychologists could erroneously conclude that an elevated RC6 score indicates that an adolescent is experiencing psychotic symptoms.

RC7 was designed to assess for negative emotional experiences related to anxiety and irritability (Archer et al., 2016). The authors listed suicidal ideation, anxiety, nightmares, problems with concentration, many specific fears, insecurity, and low self-esteem as established empirical correlates (Archer et al., 2016). As a promising finding, RC7 empirical correlates in the current study also clustered around depressive and anxious symptoms; yet, important differences emerged when examining correlates on an item level (see Table 24). For example, the current study went beyond replicating Archer et al.'s (2016) identification of anxiety as an empirical correlate. Its empirical correlates in the present study were grouped around more generalized anxiety and its consequences, as well as symptoms related to panic attacks. Of note, some differences may be attributed to the external measures used in each study. As an illustrative

example, the K-SADS-PL—used in the current study—did not assess specific fears or phobias; thus, the correlate of having many specific fears was unable to be replicated.

Table 24

Comparison of Correlations between RC7 and Selected Depressive/Anxious Symptomatology in MMPI-A-RF Clinical Correlate Studies

Disorder	Item	Veltri et al. (2009)		Forbey & Ben-Porath (2003)		Current Study	
		M	F	M	F	M	F
MDD	Depressed mood	.22	.24			.41	.22
	Anhedonia	.08	.03			.27	.17
	Suicidal ideation	.08	.24	.11	.10	.22	.29
	Self-harm	-.05	.15	.00	.19	.25	.22
	Fatigue	.11	.10			.42	.03
	Disturbance in concentration	.35	.00			.35	.19
	Psychomotor retardation	.12	.08			-.12	.26
	Worthlessness/Low Self-esteem	-.06	.00	.00	.08	.00	.07
	Hopelessness	.07	.00			.34	.10
GAD	Overanxious	.17	.04	.02	.27	.38	.29

Note: Large correlations (≥ .53) are bolded. MDD = Major Depressive Disorder; GAD = Generalized Anxiety Disorder.

Given Archer et al.'s (2016) finding of anxiety as a general RC7 empirical correlate, researchers might expect that all domains of anxiety would be at least somewhat correlated with RC7. PTSD features—no longer considered by DSM-5 in the anxiety domain—were also not related in the current dissertation. Other MMPI-A-RF clinical correlate research also failed to find correlates between PTSD and RC7 at the rigorous .35 threshold. But, lowering the threshold to .20 did produce some clinical correlates. For example, Veltri et al. (2009) documented a relationship between RC7 and flashbacks and nightmares in acute psychiatric inpatient males,

but not their female counterparts. These contrasting findings may highlight the MMPI-A-RF's difficulty in assessing PTSD or the complexity and heterogeneity of the PTSD criteria.

Differences between Higher Order and Reconstructed Clinical Scales

As discussed previously, the MMPI-A-RF, as compared to its MMPI-A predecessor, has introduced a three-tiered hierarchical structure (Archer, 2017). According to Archer et al. (2016), interpretation of the MMPI-A-RF should generally follow a top-down approach: starting with the broad Higher-Order (H-O) scales, moving to the midlevel RC scales, and ending with the narrow Specific Problems (SP) scales. In following this line of interpretation, elevated H-O scales become critically important, as they are the first area of consideration for clinicians. As such, the current dissertation deemed a systematic evaluation of the H-O scale interpretive statements to be an important topic for study.

Clinical correlates stemming from the current study offered evidence of EID's convergent validity and generally corroborated Archer et al.'s (2016) findings. According to Archer and his colleagues (2016), the Emotional/Internalizing Dysfunction (EID) scale was designed to measure overall emotional distress and dysfunction, akin to a 27/72 MMPI-A codetype. As a result, the identified empirical correlates are broadly described as "a broad range of symptoms and difficulties associated with demoralization, low positive emotions, and negative emotional experiences (e.g., low morale; depression; anxiety; feeling overwhelmed, helpless, and pessimistic)" (Archer et al., 2016, p. 74). The present study not only corroborated this description by identifying correlates clustered around depressive and anxious symptomatology but also identified generally stronger relationships among depressive symptomatology (*MDD r* = .78, items ranging from .35 to .70) than anxious symptomatology

(*GAD r* = .70 and *Panic r* = .47, items ranging from .36 to .53). Additionally, the current study identified one hypomanic feature (i.e., racing thoughts) to be moderately correlated with EID. These findings raise the possibility that EID may be more skewed towards depressive impairment, as opposed to anxious dysfunction.

Compared to the current investigation, previous research (Handel et al., 2011) demonstrated more modest discriminant validity on EID. On a scale-level, Handel and others (2011) found at least moderate relationships between EID and nearly all YSR scales, regardless of their internalizing or externalizing nature. For example, both the Internalizing (r = .64) and Externalizing (r = .43) YSR scales were significantly related to EID. Interestingly, this pattern was observed for all three H-O scales. One possible explanation for such unexpected relationships may be the general level of distress experienced by the sample. That is, it is possible that participants in Handel et al. (2011) were self-reporting more generalized distress than participants in other studies, including the current research. As one potential piece of evidence, such diffuse patterns were only evidenced on YSR scales (i.e., self-report) and not the CBCL or DBRS scales (i.e., collateral sources) included in the same study.

Strong patterns of convergent and discriminant validity were observed for THD in the present study and across the MMPI-A-RF literature. As its name suggests, the Thought Dysfunction (THD) scale's interpretation revolves around disordered thinking, predominantly as it relates to psychosis. In particular, Archer et al. (2016) explained its empirical correlates represent symptoms and difficulties related to hallucinations and delusions. Importantly, current study findings verified such findings by establishing large and moderate correlations with hallucinations and delusions, respectively. Despite general convergence across studies, relevant differences appeared when examining the magnitude of correlations between THD and

hallucinations. For example, the current studies relationships generally exceeded previously reported correlations by nearly double. As the sole exception, Veltri et al. (2009) reported a strong relationship between THD and auditory hallucinations in males, which mirrored the relationship found in the current study. One possibility for such a discrepancy involves the type of external criteria used in that structured interviews provide more detailed clinical information with symptom severity ratings. In contrast, record reviews often only provide dichotomous information (presence and absence). Unlike Veltri et al.'s (2009) emphasis on positive symptoms, the current study also included negative symptoms of psychosis, such as flat affect.

The current investigation found additional THD empirical correlates related to concentration difficulties and aggression, contrary to the MMPI-A-RF manual (Archer et al., 2016) and the broader MMPI-A-RF clinical correlate literature (Forbey & Ben-Porath; Handel et al., 2011). As one related finding, Veltri et al. (2009) identified aggressiveness as a correlate in psychiatric inpatient males. However, this correlate was only established using the comparatively lax criteria of .20. Furthermore, the authors failed to find such a relationship in females. As such, the current findings are considered preliminary and future research should focus on more detailed investigations.

Lastly, clinical correlates from the current investigation supported the Behavioral/Externalizing Dysfunction (BXD) scale's convergent validity, but provided relatively less support for its discriminant validity. Archer and others (2016, p. 76) specified BXD's empirical correlates as "a broad range of behaviors and difficulties associated with undercontrolled behavior (e.g., conduct-disordered and oppositional behaviors, alcohol or substance abuse, poor impulse control, school suspensions, running away)." In terms of convergent validity, using such a broad definition allowed all BXD empirical correlates in the

current investigation to be considered consistent. Yet, some of the PTSD-related correlates identified earlier limited its discriminant validity due to their internalizing nature. It is obvious that PTSD feature of engaging in reckless or destructive behavior as a consequence of trauma is related to under-controlled behavior and accordingly to BXD. However, the other moderate PTSD correlates of nightmares, insomnia, and detachment are not as easily understood.

The MMPI-A-RF literature appeared relatively consistent in the identification of BXD empirical correlates, but remarkably discrepant in terms of magnitude. A consistent pattern emerged in which the current study exhibited much stronger relationships than had previously been documented in the literature. For example, the current study found a very large correlation between BXD and running away (r = .67). Contrastingly, Forbey and Ben-Porath (2003) and Veltri et al. (2009) found comparatively smaller relationships, with moderate ($M\ r$ = .36) and more modest ($M\ r$ = .24) correlations, respectively. It was hypothesized that such discrepant findings may reflect differences in sample populations. The current sample included multiple adolescents detained in a juvenile detention center, a population that has been shown to exhibit increased problematic externalizing behavior (Dixon, Howie, & Starling, 2004, 2005; Tarolla, Wagner, Rabinowitz, & Tubman, 2002). In contrast, the relatively smaller relationships were documented in samples of either adolescents in a residential treatment center (Forbey & Ben-Porath, 2003) or acute psychiatric inpatients (Veltri et al., 2009). Thus, the stronger correlations may be the result of having wider variability in externalizing behavior in the current sample. To test this hypothesis directly, correlations were re-examined separated by setting (i.e., partial hospitalization versus detention). Such correlations lent support to the hypothesis by finding stronger correlations in detained adolescents than their partially hospitalized peers. Returning to the previous example of running away, detained adolescents evidenced a very large correlation (r

= .89), whereas partially hospitalized adolescents' relationship was still large but comparatively smaller (r = .54).

As an important practical consideration, it is critical to consider how the H-O scales differ from their RC scale successors in terms of clinical correlates. Given the composition of the H-O scales, sizable overlap between the clinical correlates of H-O and RC scales would be expected. The current findings largely supported this hypothesis. As summarized in Table A.2 and A.3 in the Appendix, H-O scales rarely provided unique correlates compared to their RC scale counterparts. This is an expected finding given the H-O scales' focus on broad constructs, by design. Predictably, some specificity of clinical correlates was lost when looking more broadly at the general H-O scales, as opposed to the more specific RC scales. Considering the large overlap in empirical correlates, interpreting both levels of scales may not be necessary in most cases. Thus, it may be more efficacious to focus primary interpretation on the RC scales and interpret H-O scales when they are elevated but without a corresponding RC scale elevation.

The Importance of Gender for Interpretive Statements

As detailed in the Introduction, questions have historically been raised about whether gender should be considered within interpretation of multi-scale inventories, including the MMPI family (Butcher, 2010; Butcher & Williams, 2012). Given the MMPI-A-RF's decision to use non-gendered norms, a second yet important goal of the current dissertation was to investigate the role of gender in its interpretation.

Epidemiological studies (e.g., Costello, Mustillo, Erkanli, Keeler, & Angold, 2003; Hayward & Sanborn, 2002; Merikangas et al., 2010) found gender differences in the prevalence of psychopathology in adolescents. Moreover, clinical research has suggested such gender

differences may be the most salient when examining psychopathology along the internalizing/externalizing spectrum. Consistent with the literature (Hayward & Sanborn, 2002; Merikangas et al., 2010), females in the current study tended to endorse more internalizing symptoms. Regarding RC scales, more internalizing distress manifested as higher elevations on RCd, RC1, and RC7 than their male counterparts. In fact, females (M = 66.64, SD = 13.41) averaged one standard deviation higher elevations than males (M = 52.65, SD = 13.53) on the demoralization factor. Predictably, males reported more externalizing symptomatology than females. Specifically, males exhibited higher scores on RC4 than females, producing a medium effect size (d = -0.56). Beyond gender differences on RC scale elevations, it is important to further investigate whether these scales result in different clinical correlates between males and females.

Important gender differences in clinical correlates appeared on scales RC6 and RC8. As reported, RC6 and RC8 were primarily related to psychotic symptomatology; however, additional analyses demonstrated how non-gendered interpretation obscured important differences. For example, both RC6 and RC8 were substantially ($r \geq .35$) related to psychotic symptoms in males, with no corresponding RC6 associations observed in their female peers. One explanation may lie in the distribution of scores among males and females. However, males and females exhibited comparable means and standard deviations on RC8. In contrast, females evidenced a comparatively more restricted range of RC6 scores than their male counterparts, potentially accounting for the lack of relationship among females.

This pattern of RC6 and RC8 performing accurately in males, but not females, has been previously demonstrated in the MMPI-A-RF clinical correlate literature (Veltri et al., 2009). Like the current study, Veltri and his colleagues (2009) documented (a) a stronger relationship for

males between RC8 and auditory hallucinations, but (b) a lack of substantial relationship between psychotic symptomatology and RC6 for females. In direct contrast, Forbey and Ben-Porath (2003) documented the opposite pattern: the correlation between RC8 and auditory hallucinations in females was approximately double (r = .45) than that in males (r = .23). Puzzlingly, in the same study, no significant gender differences were observed on RC6, as it was not substantially correlated with psychotic symptomatology in either gender.

One of the most prominent gender differences appeared in the MMPI-A-RF RC scales' relationship to PTSD symptomatology. A clear pattern emerged in which several RC scales (i.e., RC7 and RC8) were at least moderately correlated with trauma-related symptomatology exclusively in males (see Table 25). In complete contrast, neither RC7 nor RC8 produced a correlation exceeding .25 with PTSD symptoms in females. The gender difference in the relationship between RC8 and re-enactment of the trauma clearly exemplifies the magnitude of this difference: males produced a very large positive correlation whereas the correlation in females was negative. Moreover, several other PTSD features differed as a function of gender by over .50. It is important to consider if any other factors may have contributed to this difference. One possibility considered was that males were reporting more severe symptoms of PTSD than their female counterparts. However, a one-way ANOVA disproved this hypothesis by demonstrating comparable number of PTSD symptoms between genders. As another possibility, it was posited that more males were meeting threshold for a PTSD diagnosis than females. Again, a one-way ANOVA did not support this hypothesis. As a final consideration, as referenced above, RC8 appears to function more accurately in males. Thus, it is possible that such relationships in females exist, but RC8 is not able to accurately capture them.

Table 25

Gender Differences between RC7 and RC8 Correlations with PTSD Symptomatology in MMPI-A-RF Clinical Correlate Studies

	Veltri et al. (2009)				Forbey & Ben-Porath (2003)				Current Study			
	RC7		RC8		RC7		RC8		RC7		RC8	
	M	F	M	F	M	F	M	F	M	F	M	F
Recurrent thoughts or images									.40	.03	.44	.12
Efforts to avoid thoughts or feelings									.33	.00	**.58**	.00
Nightmares	.20	.14	.36	.01	.11	.22	.13	.27	.15	-.04	.47	-.02
Insomnia	.02	.08	.02	.11	.01	.10	.02	.29	.05	.05	.50	-.03
Irritability									.44	.02	.39	-.02
Re-enactment									.36	-.10	**.67**	-.08
Dissociative episodes									.23	.01	**.55**	.00
Distress to exposure									.10	-.16	**.55**	.00
Effort to avoid physical reminders									.11	.08	.35	.15
Inability to recall details									.31	.02	**.55**	.01
Diminished interest in activities									.21	.00	.46	-.01
Detachment									.40	.00	**.55**	.06
Restricted affect									.21	.00	.32	-.10
Sense of foreshortened future									.18	-.01	.45	.09
Difficulty concentrating									.39	-.04	.45	.12
Hypervigilance									.36	-.03	.47	.06
Exaggerated startle response									**.54**	-.05	.51	.07
Physiological reactivity									.42	.08	.51	.06
Flashbacks	.27	.06	.41	.10					.06	.01	.41	.01

	Veltri et al. (2009)				Forbey & Ben-Porath (2003)				Current Study			
	RC7		RC8		RC7		RC8		RC7		RC8	
	M	F	M	F	M	F	M	F	M	F	M	F
Negative beliefs									.24	-.05	**.55**	.00
Distorted cognitions									.06	.01	.21	.02
Negative emotional state									.25	.16	.41	-.08
Inability to experience positive emotions									.30	.25	.36	.20
Reckless or self-destructive behavior									.19	-.05	**.59**	-.03

Note: Large correlations ($\geq$ .53) are bolded. RC7 = Dysfunctional Negative Emotions; RC8 = Aberrant Experiences.

Importantly, the current investigation appears to be the first study to address PTSD correlates directly. Nonetheless, a few PTSD-relevant features were examined in prior research. For example, Veltri et al.'s (2009) record review assessed PTSD-relevant symptomatology, such as nightmares and flashbacks. Both nightmares and flashbacks were moderately correlated with RC8 in males, partially corroborating the current study. Yet, a different study by Forbey and Ben-Porath (2003) failed to find a similar relationship between nightmares and any of the scales. Additional research using a PTSD specific measure, such as the Clinician Administered PTSD Scale (CAPS; Blake et al., 1990), is clearly warranted to elucidate further these relationships.

As another key difference, RC2 appeared to function very differently by gender with respect to clinical correlates. In particular, RC2's convergent validity in the current study was moderate among females but small in males (see Table 16). Importantly, other researchers (Forbey & Ben-Porath, 2003; Handel et al., 2011) have also recognized gender differences on RC2. Handel et al. (2011) documented moderate relationships between RC2 and the YSR's Withdrawn/Depressed (r = .43) and Internalizing (r = .35) scales among females, but consistently weak relationships (rs < .15) in males. A large disparity was also observed in Forbey and Ben-Porath (2003), with moderate correlations being observed on depression-relevant ACDF scales (i.e., ACDF Depression and ACDF Self-Esteem) in females (M = .42), but not males (M = .16). Thus, when looking across the prior literature, a consistent pattern emerges in which RC2's convergent validity is poor among males. Consequently, clinicians may wish to exercise caution when interpreting RC2 in that population and consult RCd for additional interpretation of depressive symptoms in males.

Clearly, gender differences represent an interpretive concern. In particular, the current study demonstrated differences in terms of scale elevations and potential differences in what

those elevations may indicate. Pragmatically, it is important for clinicians to reflect on how gender should be considered when interpreting the MMPI-A-RF. One gender-sensitive approach would be to have interpretation informed by gender-specific empirical correlates. Practitioners may be interested in how such an approach would look in practice. The following section focuses on this point, among other practical considerations.

Clinical Implications

Clinically, the MMPI-A-RF provides important information that may affect diagnostic conceptualization, as it assesses a wide range of psychopathology and personality characteristics (Archer et al., 2016). Given its recent publication date, the current study sought to thoroughly examine its utility as an emerging assessment measure for adolescents. Importantly, the current study mostly lent support to the utility of the RC scales in identifying distress and dysfunction along the internalizing/externalizing spectrum. Yet, additional findings also raised questions about (a) the redundancy of the H-O scales with RC scales (b) limited empirical correlates for some RC scales and (c) gender differences related to interpretative statements.

Top-Down Approach to MMPI-A-RF Empirical Correlates

Practitioners have received a recommendation from the MMPI-A-RF authors (Archer et al., 2016; Archer, 2017) to follow a top-down approach to its interpretation: beginning with the broad H-O scales, then assessing individual RC scales, and finally examining the more narrow SP scales. On the positive side, all three H-O scales demonstrated good patterns of convergent and discriminant validity on both a scale and item level. The H-O scales also produced patterns of clinical correlates with conceptually relevant K-SADS-PL items. Importantly, relativity little

overlap within scales increases their interpretatively usefulness. As previously noted, EID, THD, and BXD were most related to internalizing distress, psychotic symptomatology, and externalizing behaviors, respectively. As a result, H-O scales serve as a good indicator to clinicians about the potential distress or dysfunction adolescents may be experiencing on a broad internalizing versus externalizing domain level.

As could be anticipated, the MMPI-A-RF faced more difficulties identifying psychopathology beyond the broad internalizing/externalizing domain. Keeping in mind the top-down approach recommended by Archer et al. (2016), the RC scales should produce more refined interpretations than the H-O scales. Yet, patterns of convergent and discriminant validity, as well as clinical correlates, generally mirrored those of the H-O scales. Such redundancy between H-O and RC scales limits their combined interpretive utility. As a salient example, consider the redundancy between BXD and RC4, in which 75% of items on RC4 are also included on BXD. Unfortunately, clinical correlates of the two scales are virtually identical as a result. Such significant overlap, does call into question any added advantage of using the recommended top-down approach to interpretation. Considering how some specificity in terms of clinical correlates is lost when utilizing the broader H-O scales, it may be more practical for clinicians to focus their interpretation on RC scales. Such an approach would reduce redundancy and improve efficiency. As one potential concern, that may limit interpretation for a small number of individuals who had an elevated H-O scale without a corresponding elevation on the relevant RC scale. Continuing with the BXD/RC4 example, one adolescent would have been excluded from interpretation. Thus, the best strategy may be to interpret H-O scales in cases in which H-O scales are elevated without corresponding RC scale elevations.

Interestingly, the RCd scale was purposefully implemented in the redesigned MMPI-A-RF as a means of reducing the noted redundancy among the basic scales of its predecessor. Archer (2017) aptly summarized the rationale:

> The central objective in the development of the MMPI-A-RF was to improve on the discriminate validity achievable by the MMPI-A by reducing the ubiquitous and confounding influence of the demoralization factor commonly found in most personality measures… While scales heavily influenced by the Demoralization factor might be expected to show strong evidence of convergent validity (i.e., high correlations with predicted external criteria), such scales typically suffer from relatively poor specificity or discriminate validity (i.e., the ability to discriminate between various forms of psychopathology)… The MMPI-A-RF seeks to reduce the redundancy found among MMPI-A scales by isolating the demoralization factor and reducing its influence on the 'seed' or 'core' components of the MMPI-A-RF scales. (p. 328)

Encouragingly, current results generally documented improved discriminant validity among RC scales. Moreover, as envisioned by Archer et al. (2016), RCd was confirmed as an indicator of distress, particularly in the internalizing realm. That is, it was robustly related to affective and anxious distress. As a potentially unintended consequence, RCd appeared to be a better indicator of depressive symptoms than RC2. Thus, while arguably improved in the MMPI-A-RF, current study findings demonstrated that discriminant validity remains a concern.

Some poor discriminant validity among MMPI-A-RF scales affected its use to detect more specific psychopathology. In the internalizing realm, both RCd and RC7 were largely related to both affective and anxious psychopathology. In fact, elevated RC7 scores were predictive of both Major Depressive Disorder and Generalized Anxiety Disorder. While such overlap may in part be expected due to recognized comorbidity between depression and anxiety (r = .67 in the current study), it muddies the ability to make more nuanced interpretation of the scales. On the other end of the spectrum, RC4 was correlated with nearly all externalizing disorders assessed in the current study, and particularly predictive of Conduct Disorder and Substance Use Disorder. In both RC7 and RC4, established cut scores functioned best for ruling

out certain psychopathology with low scores than ruling it in with high scores. Thus, as a positive finding, practitioners can feel confident that adolescents scoring below established cut scores on RC7 and RC4 will be unlikely to meet clinical criteria for relevant DSM disorders.

Strengths and Limitations of Empirically Based Interpretations for MMPI-A-RF RC Scales

The current investigation builds on the initial work of Archer and colleagues (2016) concerning their inclusion of interpretive statements for each MMPI-A-RF scale. According to Archer et al., such interpretive statements are based off of a combination of previous empirical correlate research (e.g., Handel et al., 2011; Veltri et al., 2009) and an analysis of item content. As a particularly encouraging finding, the current dissertation replicated and/or expanded previously identified empirical correlates for the majority of MMPI-A-RF RC scales. For such correlates that are replicated across several studies, clinicians can feel confident in including them in their assessment reports. That is not to say that only empirical correlates replicated in the current study should be included. Instead, practitioners might choose to give greater weight to those statements with strong empirical support.

On the other hand, there were a few MMPI-A-RF RC scales with little to no empirical correlates in the current study, namely RC9. To improve empirical support for such scales, scholars should focus on its psychometric properties. Collectively, MMPI-A-RF empirical correlate studies consistently identified RC9 as producing the lowest reliability coefficients. Poor homogeneity of scale items in turn affects scale interpretability. As previously mentioned, excluding RC9 from interpretation completely would be a far too conservative approach. Instead, practitioners may rely more heavily on item content for interpretation.

Gender Differences and MMPI-A-RF Empirical Correlates

Patterns of differences between genders in the current dissertation highlight the importance of considering gender in interpretation of the MMPI-A-RF. As discussed in previous sections, the use of gender-specific versus non-gendered norms has historically been a source of contention among scholars (Ben-Porath & Flens, 2012; Butcher & Williams, 2012). With the MMPI-2-RF and MMPI-A-RF's utilization of non-gendered norms, the gendered norms debate has resurfaced. Without entering the debate, the present study demonstrated important differences in MMPI-A-RF profiles between genders. In fact, average scale elevations between genders differed by nearly 10T on some scales. Beyond scale elevations, the current dissertation further documented how gender differences may impact scale interpretation. Nearly half of the MMPI-A-RF RC scales exhibited drastically different clinical correlates when separated by gender. That is, even similar elevations in terms of magnitude on any of these four scales would still be indicative of different clinical presentations based on gender.

Considered together, current results indicate that gender clearly needs to be considered during interpretation of the MMPI-A-RF. One approach would be for clinicians to focus on empirically based gender-specific interpretation, regardless of gendered versus non-gendered norms. That is, clinicians would look to gender-specific clinical correlates to inform interpretation. In his notable work on the interpretation of the MMPI-2/MMPI-2-RF, Greene (2011) published empirical correlates for scales separated by gender. This practice, in turn, allowed practitioners to more easily take gender into account when making interpretations. Building off of Greene (2011), a similar methodology with the MMPI-A-RF appears to be a reasonable approach to considering gender, without necessitating gendered norms.

Limitations and Future Directions

Limitations

The current study had several important methodological limitations. First, because of challenges to recruitment, this sample was limited to 66 adolescents. While a broad array of psychopathology was observed in the current sample, several important clinical presentations were not adequately represented. As a result, it is entirely possible that additional correlates between MMPI-A-RF RC scales and symptomatology exist, but low representativeness in the current sample disallowed them to be evident. Importantly, those disorders not well represented in the current study are relatively rare in an adolescent populations overall; namely, manic and psychotic symptomatology. A logical next step would be to explore the relationships between MMPI-A-RF RC scales and these symptomatology in samples in which these diagnoses are over-sampled. For example, conducting a similar study in a more specialized site dedicated solely to treating adolescents with severe mental disorders (i.e., severe mood and/or psychotic disorders) may allow for more clinical variability.

At present, studies examining the MMPI-A-RF have typically employed a majority of European Americans (Archer et al., 2016), with the current study being no exception. Like previous research, the limited ethnic diversity potentially reduces the generalizability of its findings. Furthermore, previous empirical research (Cashel et al., 1998) on the MMPI-A-RF's predecessor, the MMPI-A, found significant ethnic differences. Consequently, a priority for future MMPI-A-RF research should be to examine its reliability and validity among more ethnically diverse samples.

As a final limitation, consistent with the MMPI-A-RF extant literature, the current study examined gender as a dichotomous variable. However, as emphasized by Zeanah and Myint

(2017), parallel to shifting cultural norms, psychological research is increasingly considering gender nonconformity among children and adolescents. To date, the impact of gender nonconformity has not been considered in the context of the MMPI-A-RF. Given consistent findings that these children and adolescents are at increased risk for psychopathology (Aitken, VanderLaan, Wasserman, Stojanovski, & Zucker, 2016; Zeanah & Myint, 2017), it is critical to examine the utility of the MMPI-A-RF in this population.

Future Directions

An obvious area of future research should center on continued empirical examination of the MMPI-A-RF. One interesting avenue of study would involve the test-retest reliability of the MMPI-A-RF RC scales among specific adolescent populations. While Archer et al. (2016) provided test-retest reliability statistics for all MMPI-A-RF scales in the normative sample, it would be important to examine how such reliabilities differ in clinical and forensic samples, where populations with more psychological impairment may demonstrate marked variability. Moreover, researchers could examine changes MMPI-A-RF RC scales as outcome variables for mental health treatment.

The current dissertation focused on the clinical correlates of the MMPI-A-RF substantive scales, but continued empirical investigation of the remaining MMPI-A-RF scales is critical. Particularly, the Specific Problems (SP) scales and the Personality Psychopathology Five (PSY-5) scales of the MMPI-A-RF were not addressed. Following the approach of the current study, investigation of the SP and PSY-5 scales utilizing a structured interview would be a useful direction for future research. For example, the Child Assessment Schedule (CAS; Hodges, Kline, Stern, Cytryn, & McKnew, 1982) is a brief structured interview intended to assess an

individual's domains of functioning, such as family and school. Adding to the methodological rigor of the approach, the CAS can be administered to collateral sources, such as a parent or a caregiver. Including information from collateral sources would be an invaluable addition given discrepancies in the reporting of psychopathology between adolescents and collateral sources can occur (De Los Reyes & Kazdin, 2005).

Another logical next step is empirical examination of the revised validity scales and their ability to detect response distortion. As documented by Archer (2017), the F-r scale represents a substantive change from its predecessor. As part of the revision, the inclusion criteria for F-r items were relaxed to include items with endorsement frequencies below 15% in the normative sample and 20% in the development sample. Such endorsement rates are considerably higher than other infrequency scales employing a rare symptom strategy. Thus, it is critical to examine the efficacy of the F-r in light of the comparatively lax standard. Furthermore, several MMPI-A validity indicators, such as F1 and F2, were surprisingly not included on the MMPI-A-RF. Researchers may explore how those validity scales would perform on the MMPI-A-RF. Studies exploring both overreporting and underreporting in a variety of research designs (i.e., simulation designs vs. known group comparisons) present an exciting avenue for future research.

Final Considerations

Together with the existing MMPI-A-RF literature, the current investigation generally supports the utility of the MMPI-A-RF as an emerging measure in the realm of adolescent assessment. Generally, the present study bolstered support for the construct validity of the RC scales through the identification of clinically relevant, scale and item-level correlates. In particular, the RCd and RC4 evidenced especially strong convergent and discriminant validity.

As a result, the current study highlighted the MMPI-A-RF's efficacy in detecting psychopathology and dysfunction along the broad externalizing and internalizing spectrum.

Nevertheless, the current dissertation also identified some key areas for improvement. In agreement with Archer (2017), its findings caution against using the results to determine specific diagnoses. Instead, the MMPI-A-RF may function best as part of a multifaceted approach to assessment. That is, practitioners may wish to use the MMPI-A-RF as an important starting place in assessing adolescents' mental health functioning by determining where the adolescent's distress lies broadly and ruling out specific symptomatology. Then, clinicians can follow up on MMPI-A-RF results through a series of test batteries purposely designed to assess specific psychopathology and distinguish among relevant disorders. As documented, continued work on empirical correlates separated by gender represent an important area for research to ensure accurate empirically based, gender-specific interpretation of the MMPI-A-RF.

APPENDIX

SUPPLEMENTARY TABLES

Table A.1.

Differences on RC7 and RC4 Scales between Adolescents With and Without Common Diagnoses

	Diagnosis		No Diagnosis				
	M	*SD*	*M*	*SD*	*F*	*p*	*d*
RC7							
MDD	58.00	10.76	46.82	7.90	23.33	<.001	1.19
GAD	57.55	11.13	47.54	8.31	17.37	<.001	1.03
RC4							
CD	69.80	13.43	48.43	9.86	45.23	<.001	1.98
SUD	67.00	13.91	46.98	8.40	46.76	<.001	1.99

Note. MDD = Major Depressive Disorder, GAD = Generalized Anxiety Disorder, CD = Conduct Disorder, SUD = Substance Abuse Disorder.

Table A.2

Substantial Correlations (≥ .35) of MMPI-A-RF Internalizing H-O and RC Scales with Selected K-SADS-PL Internalizing Item-level Correlates

K-SADS-PL		MMPI-A-RF scales							
Section	K-SADS-PL item stem	EID	RCd	RC1	RC2	RC7	THD	RC6	RC8
MDD	Depressed mood	**.69**	**.65**	.43		.41			
	Anhedonia	**.63**	**.64**		.39				
	Recurrent thoughts of death	**.62**	**.61**	.35		.36			
	Suicidal ideation	**.53**	**.53**			.36			
	Self harm	.47	.46						
	Lack of reactivity	**.70**	**.68**	.45	.39	.41			
	Quality of dysphoria	.43	.50						
	Sleep disturbance - initial			.36		.35			
	Non-restorative sleep		.38	.39					
	Fatigue	**.53**	.51	.42					
	Disturbance in concentration	.35					.36	.36	
	Decreased Appetite	.50	.45			.38			
	Weight loss	.42				.36			
	Worthlessness	.47	.45						
	Hopelessness	.49	.49		.40				
Bipolar Disorder	Racing thoughts	.35							
	Influence of alcohol or drugs	-.35							
Psychosis	Hallucinations						**.53**	.35	**.61**
	Delusions						.43		
	Flat affect						.41		.46

K-SADS-PL		MMPI-A-RF scales							
Section	K-SADS-PL item stem	EID	RCd	RC1	RC2	RC7	THD	RC6	RC8
	Inappropriate affect						.50	.38	**.53**
	Incoherence						.52	.37	**.57**
	Loosening of associations						.52	.37	**.57**
	Catatonic behavior						.52	.37	**.57**
Panic Disorder	Panic attacks	.50	.47	.48		.42			
	Fear of another attacks	.49	.47	.38		.42			
	Agoraphobia	.53	.49	.42		.45			
	Maladapt. changes to avoid attacks					.39			
GAD	Excessive worry	.51	.44			.42			
	Overconcern with competence	.36				.40			
	Need for reassurance					.36			
	Ability to control worries	.43	.41			.41			
	Muscle tension	.44	.48	.47		.41			
	Restlessness	.45	.38	.35		.51			
	Easily fatigued	.47	.39	.35		.42			
	Feeling keyed up	.42	.37						
	Difficulty concentrating	.43	.43						
	Trouble falling or staying asleep		.35						

Note: Large correlations ($\geq$ .53) are bolded. EID = Emotional/Internalizing Dysfunction; THD = Thought Dysfunction; RCd = Demoralization; RC1 = Somatic Complaints; RC2 = Low Positive Emotions; RC3 = Cynicism; RC6 = Ideas of Persecution; RC7 = Dysfunctional Negative Emotions; RC8 = Aberrant Experiences.

Table A.3

Substantial Correlations (≥ .35) of MMPI-A-BXD and RC4 Scales with Selected K-SADS-PL Externalizing Item-level Correlates

K-SADS-PL		MMPI-A_RF Scale	
Section	K-SADS-PL item stem	BXD	RC4
ADHD	Impulsivity	.38	
	Engages in dangerous activities	.43	.39
ODD	Loses temper	.51	.40
	Argues a lot with adults	.39	
	Disobeys rules	**.61**	**.65**
	Easily annoyed	.37	.37
	Spiteful or vindictive	.46	.47
	Uses bad language	.46	.49
	Annoys people on purpose	.45	.46
Conduct Disorder	Lies	.45	**.53**
	Truant	.44	**.55**
	Initiates physical fights	**.60**	**.58**
	Bullies, threatens, or intimidates	.45	.37
	Nonaggressive stealing	**.57**	**.59**
	Vandalism	**.59**	**.58**
	Breaking and entering	.52	.48
	Aggressive stealing	**.58**	**.55**
	Often stays out at night	**.62**	**.60**
	Ran away overnight	**.67**	**.67**
	Use of a weapon	**.54**	**.53**
	Physical cruelty to others	.51	.52
Alcohol Use	Frequency	.52	**.55**
	Concern from others	**.56**	**.59**
	Drinks more than planned	**.53**	**.53**
	Negative physical consequences	.52	**.57**
	Dangerous behavior	**.63**	**.64**
	Negative psych. consequences	**.58**	**.61**
	Negative occupational consequences	**.62**	**.66**
	Negative social consequences	**.64**	**.68**

K-SADS-PL		MMPI-A_RF Scale	
Section	K-SADS-PL item stem	BXD	RC4
	Legal consequences	**.54**	**.57**
	Failure to fulfill major obligations	**.57**	**.58**
	Important activities given up	**.71**	**.74**
	Time consuming	**.55**	**.60**
	Tolerance	**.57**	**.59**
	Tried to quit	**.58**	**.60**
	Withdrawal	**.57**	**.60**
	Alcohol consumed to relieve withdrawal	**.60**	**.61**
Substance Use	Frequency	**.59**	**.72**
	Uses more than planned	**.53**	**.60**
	Negative physical consequences	**.61**	**.68**
	Dangerous behavior	**.67**	**.76**
	Negative psych. consequences	**.56**	**.63**
	Negative occupational consequences	**.61**	**.75**
	Negative social consequences	**.56**	**.66**
	Legal consequences	**.66**	**.75**
	Failure to fulfill major obligations	**.56**	**.68**
	Important activities given up	**.61**	**.71**
	Time consuming	**.61**	**.69**
	Tolerance	**.57**	**.70**
	Tried to quit	.50	**.61**
	Withdrawal	**.56**	**.66**
	Drugs consumed to relieve withdrawal	**.58**	**.67**

Note: Large correlations ($\geq$ .53) are bolded. Unique correlates are underlined. BXD = Behavioral/Externalizing Dysfunction; RC4 = Antisocial Behavior.

Table A.4

MMPI-A-RF RC Scale Alphas Separated by Gender

Scale	Total	Males	Females
RCd	.92	.91	.88
RC1	.85	.85	.84
RC2	.69	.71	.67

Scale	Total	Males	Females
RC3	.63	.52	.66
RC4	.88	.86	.89
RC6	.72	.78	.64
RC7	.66	.69	.56
RC8	.57	.61	.56
RC9	.47	.51	.47

Note: RCd = Demoralization; RC1 = Somatic Complaints; RC2 = Low Positive Emotions; RC3 = Cynicism; RC4 = Antisocial Behavior; RC6 = Ideas of Persecution; RC7 = Dysfunctional Negative Emotions; RC8 = Aberrant Experiences; RC9 = Hypomanic Activation.

Table A.5

Comparison of Correlations between RC4 and Externalizing Symptomatology in MMPI-A-RF Clinical Correlate Studies

	Veltri et al. (2009)		Forbey & Ben-Porath (2003)		Current Study	
	M	F	M	F	M	F
History of Criminal Charges	.48	.40	-.10	.42		
Aggressiveness	.26	.15	.20	.47		
Impulsive	-.09	.11	.09	.34		
Running Away	.17	.31	.33	.46	.69	.75
Fighting	.29	.11	.02	.31	.64	.46
Running Away	.23	.32	.32	.52		
Assault	.08	.01	.10	.22	.64	.33
Fire-Setting	-.01	.00	.02	.26	.10	.30
Stealing <$50	.18	.23	.12	.16		
Stealing >$50	-.07	.09	.13	.29		
Oppositional Behavior	.30	.44				
Poor Judgment	.03	.08				
Anger	.10	.14			.11	.31
Truancy	.12	.18			.61	.57
Suspension	.39	.29				
Peers Bad Influence	.34	.24				
Lying	.00	.20			.51	.50
History of Substance Abuse			.35	.54		

	Veltri et al. (2009)		Forbey & Ben-Porath (2003)		Current Study	
	M	F	M	F	M	F
Intermittent Rage			.09	.07		
Shoplifting			.11	.18		
Drug Dealing			.15	.30		
Sexual Acting Out			-.04	.15		
Threatened with Weapon			.05	.23	.64	.36
Animal Abuse			.08	.15	.10	.00
Vandalism			.12	.14	.61	.58
Sexual Assault			.01	.04	.00	.09
Threatened Verbal			.02	.19	.25	.48
Property Offense			.20	.41		
Status Offense			.04	.42		

REFERENCES

Achenbach, T., & Edelbrock, G. (1983). *Manual for the Child Behavior Checklist and Revised Child Behavior Profile*. Burlington: University of Vermont Department of Psychiatry.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms and Profiles*. Burlington: University of Vermont, Research Center for Children, Youth, & Families.

Aitken, M., VanderLaan, D.P., Wasserman, L., Stojanovski, S., & Zucker, K.J. (2016). Self-harm and suicidality in children referred for gender dysphoria. *Journal of the American Academy of Child and Adolescent Psychiatry, 55,* 513–520.

Alperin, J. J., Archer, R. P., & Coates, G. D. (1996). Development and effects of an MMPI-A K-correction procedure. *Journal of Personality Assessment, 67*(1), 155-168.

Ambrosini, P. J., Metz, C., Prabuki, K., & Lee, J. C. (1989). Video tape reliability of the third revised edition of the K-SADS-III-R. *Journal of the American Academy of Child and Adolescent Psychiatry, 28,* 723-728.

Archer, R. P. (1984). Use of the MMPI with adolescents: A review of salient issues. *Clinical Psychology Review, 4*, 241-251. doi:10.1016/0272-7358(84)90002-3

Archer, R. P. (1992). *MMPI-A: Assessing adolescent psychopathology*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Archer, R. P. (1997). Future directions for the MMPI-A: Research and clinical issues. *Journal of Personality Assessment, 68*(1), 95-109.

Archer, R. P. (2005). *MMPI-A: Assessing adolescent psychopathology* (3rd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Archer, R. P. (2017). *Assessing Adolescent Psychopathology: MMPI-A/MMPI-A-RF* (4th ed.). New York, NY: Routledge/Taylor & Francis Group.

Archer, R. P., & Elkins, D. E. (1999). Identification of random responding on the MMPI-A. *Journal of Personality Assessment, 61,* 547-556. doi:10.1207/S15327752JPA7303_8

Archer, R. P., Griffin, R., & Aiduk, R. (1995). MMPI-2 clinical correlates for ten common codes. *Journal of Personality Assessment, 65*, 391-407.

Archer, R. P., Handel, R. W., Ben-Porath, Y. S., & Tellegen, A. (2016). *MMPI-A-RF (Minnesota Multiphasic Personality Inventory-Adolescent-Restructured Form) administration, scoring, interpretation, and technical manual*. Minneapolis: University of Minnesota Press.

Archer, R. P., Handel, R. W., & Lynch, K. D. (2001). The effectiveness of MMPI-A items in discriminating between normative and clinical samples. *Journal of Personality Assessment, 77*(3), 420-435. doi:10.1207/S15327752JPA7703_04

Archer, R. P., Pancoast, D. L., & Klinefelter, D. (1989). A comparison of MMPI code types produced by traditional and recent adolescent norms. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1*(1), 23-29. doi:10.10371040-3590.1.1.23

Archer, R. P., & Newsom, C. R. (2000). Psychological test usage with adolescent clients: Survey update. *Assessment*, 7(3), 227-235. doi:10.1177/107319110000700303

Archer, R. P., Simonds-Bisbee, E. C., Spiegel, D. R., Handel, R. W., & Elkins, D. E.

(2010). Validity of the Massachusetts Youth Screening Instrument-2 (MAYSI-2) scales in juvenile justice settings. *Journal of Personality Assessment, 92*, 337-348. doi:10.1080/00223891.2010.482009

Archer, R. P., White, J. L., & Orvin, G. H. (1979). MMPI characteristics and correlates among adolescent psychiatric inpatients. *Journal of Clinical Psychology, 35*, 498-504.

Avenevoli, S., Swendsen, J., He, J.P., Burstein, M., & Merikangas, R. K. (2015). Major Depression in the National Comorbidity Survey-Adolescent Supplement: Prevalence, correlates, and treatment. *Journal of the American Academy of Child and Adoelscent Psychiatry, 54*(1), 37-44. doi:10.1016/j.jaac.2014.10.010

Baer, R. A., Kroll, L. S., Rinaldo, J., & Ballenger, J. (1999). Detecting and discriminating between random responding and overreporting on the MMPI-A. *Journal of Personality Assessment, 72*, 308-320.

Barkley, R. A., & Murphy, K. R. (1998). *Attention deficit hyperactivity disorder: A clinical workbook* (2nd ed.). New York, NY: Guilford.

Beck, A., Ward, C., Mendelson, M., Muck, M., & Erbaugh, J. (1961). An inventory of measuring depression. *Archives of General Psychiatry, 4*, 561-571.

Ben-Porath, Y.S. (2012). *Interpreting the MMPI-2-RF*. Minneapolis: University of Minnesota Press.

Ben-Porath, Y. S., & Flens, J. R. (2012). Butcher and Williams's (This Issue) critique of the MMPI-2-RF is slanted and misleading. *Journal of Child Custody, 9*, 223-232. doi:10.1080/15379418.2012.748605

Ben-Porath, Y. S., & Tellegen, A. (2008/2011). *MMPI-2RF: Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.

Blake, D., Weathers, F., Nagy, L., Kaloupek, D., Klauminzer, G., Charney, D., & Keane, T. (1990). *Clinician-Administered PTSD Scale (CAPS)*. National Center for Post-Traumatic Stress Disorder, Behavioral Science Division Boston-VA, Boston, MA.

Burstein, M., Beesdo-Baum, K., He, J. P., & Merikangas, K. R. (2014). Threshold and subthreshold generalized anxiety disorder among US adolescents: Prevalence, sociodemographic, and clinical characteristics. *Psychological Medicine, 44*(11), 2351-2362. doi:10.1017/S0033291713002997

Butcher, J. N. (2010). Personality assessment from the 19th to the early 21st century: Past achievements and contemporary challenges. Annual Review of Clinical Psychology, 6, 1–20.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, G. W., & Kaemmer, B. (2001). *MMPI-2: Manual for administration and scoring*. Minneapolis: University of Minnesota Press.

Butcher, J. N., & Williams, C. L. (2012). Problems with using the MMPI-2-RF in forensic evaluations: A clarification to Ellis. *Journal of Child Custody: Research, Issues, and Practices, 9*, 217-222. doi:10.1080/15379418.2012.748347

Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R. P., Tellegan, A., Ben-Porath, Y. S., & Kaemmer, B. (1992). *Minnesota Multiphasic Personality Inventory-Adolescent*. Minneapolis: University of Minnesota Press.

Cantwell, D. P., Lewinsohn, P. M., Rohde, P., & Seeley, J. R. (1997). Correspondence between adolescent report and parent report of psychiatric diagnostic data. *Journal of the American Academy of Child and Adolescent Psychiatry, 36*(5), 610-619. doi:10.1097/00004583-199705000-00011

Capwell, D. F. (1945a). Personality patterns of adolescent girls: I. Girls who show improvement in IQ. *Journal of Applied Psychology, 29*, 212-228. doi:10.1037/h0062853

Capwell, D. F. (1945b). Personality patterns of adolescent girls: II. Delinquents and non-delinquents. *Journal of Applied Psychology, 29*, 212-228. doi:10.1037/h0054701

Cashel, M. L., Rogers, R., Sewell, K. W., & Holliman, N. B. (1998). Preliminary validation of the MMPI-A for a male delinquent sample: An investigation of clinical correlates and discriminant validity. *Journal of Personality Assessment, 7*(1), 49-69.

Cauffman, E., Feldman, S. S., Waterman, J., & Steiner, H. (1998). Posttraumatic stress disorder among female juvenile offenders. *Journal of the American Academy of Child and Adolescent Psychiatry, 37*(11), 1209-1216. doi:10.1097/00004583-199811000-000022

Cauffman, E., & Steinberg, L. (2000). (Im)maturity of judgment in adolescence: why adolescents may be less culpable than adults. *Behavioral Sciences & the Law, 18*(6), 741-760. doi:10.1002/bsl.416

Chase, T. V., Chaffin, S., & Morrison, S. D. (1975). False positive adolescent MMPI profiles. *Adolescence, 40*, 507-519.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309-319.

Cocozza, J. J., & Skowyra, K. R. (2000). Youth with mental health disorders: Issues and emerging responses. *Juvenile Justice, 7*(1), 3-13.

Colligan, R. C., & Offord, K. P. (1989). The aging MMPI: Contemporary norms for contemporary teenagers. *Mayo Clinic Proceedings, 64*, 3-27. doi:10.1016/S0025-6196(12)65299-9

Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry, 60*(8), 837-844.

Craig, R. J. (2003). Assessing personality and psychopathology with interviews. In J. R. Graham, J. A. Naglieri, & I. B. Weiner (Eds.), *Handbook of psychology: Assessment psychology* (Vol. 10, p. 487-508). Hoboken, NJ: Wiley.

Dahlstrom, W. G., Welsh, W. G., & Dahlstrom, L. E. (1972). *An MMPI handbook: Vol. 1 Clinical interpretation* (rev. ed.). Minneapolis: University of Minnesota Press.

Davis, T. C., Michielutte, R., Askov, E. N., Williams, M. V., & Weiss, B. D. (1998). Practical assessment of adult literacy in health care. *Health Education & Behavior, 25*(5), 613-624. doi:10.1177/109019819802500508

De Los Reyes, A., Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131*(4), 483-509. doi:10.1037/0033-2909.131.4.483

Derogatis, L. R. (1983). *SCL-90-R: Administration, scoring, and procedures manual*. Towson, MD: Clinical Psychometric Research.

Dierkhising, C. B., Ko, S. J., Woods-Jaeger, B., Briggs, E. C., Lee, R., & Pynoos, R. S. (2013). Trauma histories among justice-involved youth: Findings from the National Child Traumatic Stress Network. *European Journal of Psychotraumatology, 4*(1), 1-12. doi:10.3402/ejpt.v4i0.20274

Dixon, A., Howie, P., & Starling, J. (2004). Psychopathology in female juvenile offenders. *Journal of Child Psychology and Psychiatry, 45*(6), 1150-1158. doi:10.1111/j.1469-7610.2004.00307.x

Dixon, A., Howie, P., & Starling, J. (2005). Trauma exposure, posttraumatic stress, and psychiatric comorbidity in female juvenile offenders. *Journal of the American Academy of Child & Adolescent Psychiatry, 44*(8), 798-806. doi:10.1097/01.chi.0000164590.48318.9c

Ehrenworth, N. V., & Archer, R. P. (1985). A comparison of clinical accuracy ratings of interpretive approaches for adolescent MMPI responses. *Journal of Personality Assessment, 49*, 413-421. doi:10.1207/s15327752jpa4904_9

Fazel, S., Doll, H., & Langstrom, N. (2008). Mental disorders among adolescents in juvenile correctional facilities: A systematic review and metaregression analysis of 25 surveys. *Journal of the American Academy of Child and Adolescent Psychiatry, 47*(9), 1010-1019. doi:10.1097/CHI.0b013e31817eecf3

Fontaine, J. L., Archer, R. P., Elkins, D. E., & Johansen, J. (2001). The effects of MMPI-A *T*-score elevation on classification accuracy for normal and clinical adolescent samples. *Journal of Personality Assessment, 76*(2), 264-281. doi:10.1207/S15327752JPA7602_09

Forbey, J. D., & Ben-Porath, Y. S. (2003). Incremental validity of the MMPI-A content scales in a residential treatment facility. *Assessment, 10*(2), 191-202. doi:10.1177/1073191103010002010

Forbey, J. D., Ben-Porath, Y. S., & Davis, D. L. (2000). A comparison of sexually abused and non-sexually abused adolescents in a clinical treatment facility using the MMPI-A. *Child Abuse and Neglect, 24*, 557-568.

Gottesman, I., Hanson, D. R., Kroeker, T. A., & Briggs, P. F. (1987). New MMPI normative data and power-transformed T-score tables for the Hathaway-Monachesi Minnesota Cohort of 14,019 15-year-olds and 3,674 18-year-olds. In R. P. Archer (Ed.), *Using the MMPI with adolescents* (p. 241-292). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Graham, J. R., Ben-Porath, Y. S., & McNulty, J. L. (1999). *MMPI-2 correlates for outpatient community health settings*. Minneapolis: University of Minnesota Press.

Greene, R. L. (2011). *The MMPI-2/MMPI-2-RF: An interpretive manual* (3rd ed.). Needham Heights, MA: Allyn & Bacon.

Grisso, T. (1999). Juvenile offenders and mental illness. *Psychiatry, Psychology, and Law, 6*(2), 143-151.

Grisso, T., & Barnum, R. (2006). *Massachusetts Youth Screening Instrument-Version 2: User's manual and technical report* (Rev. ed.). Sarasota, FL: Professional Resource Press.

Groth-Marnat, G. (2009). *Handbook of psychological assessment - 5th edition*. John Wiley & Sons, Hoboken: New Jersey.

Gynther, M. D., Altman, H., & Sletten, I. W. (1978). Replicated correlates of MMPI two-point code types: The Missouri actuarial system. *Journal of Clinical Psychology, 29,* 263-289.

Gynther, M. D., & Shimkunas, A. M. (1966). Age and MMPI performance. *Journal of Consulting Psychology, 30*(2), 118-121.

Hand, C. G. (2005). *The classification accuracy of the MMPI-A: Effects of modifying the normative sample*. Unpublished doctoral dissertation, Virginia Consortium Program in Clinical Psychology.

Handel, R. W., Archer, R. P., Elkins, D. E., Mason, J. A., & Simonds-Bisbee, E. C. (2011). Psychometric properties of the Minnesota Multiphasic Personality Inventory-Adolescent (MMPI-A) clinical, content, and supplementary scales in a forensic sample. *Journal of Personality Assessment, 93*(6), 566-581. doi:10.1080/00223891.2011.608752

Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory (Rev. ed.)*. Minneapolis: University of Minnesota Press.

Hathaway, S. R., & Monachesi, E. D. (1963). *Adolescent personality and behavior: MMPI patterns of normal, delinquent, dropout, and other outcomes*. Minneapolis: University of Minnesota Press.

Hayward, C., & Sanborn, K. (2002). Puberty and the emergence of gender differences in psychopathology. *Journal of Adolescent Health, 30* (4SUPPL.1), 49-58. doi:10.1016/S1054-139X(02)00336-1

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development, 34*, 177-189.

Hilts, D., & Moore, J. M. (2003). Normal range MMPI-A profiles among psychiatric inpatients. *Assessment, 10*(3), 266-272. doi:10.1177/1073191103255494

Hodges, K., Kline, J., Stern, L., Cytryn, L., & McKnew, D. (1982). The development of a Child Assessment Interview for research and clinical use. *Journal of Abnormal and Child Psychology, 10*, 173-189.

Jensen, P. S., Rubio-Stipec, M., Canino, G., Bird, H. R., Dulcan, M. K., Schwab-Stone, M. E., & Lahey, B. B. (1999). Parent and child contributions to diagnosis of mental disorder: Are both informants always necessary? *Journal of the American Academy of Child and Adolescent Psychiatry, 38*(12), 1569-1579. doi:10.1097/00004583-199912000-00019

Jewell, J., Handwerk, M., Almquist, J., & Lucas, C. (2004). Comparing the validity of clinician-generated diagnosis of conduct disorder to the Diagnostic Interview Schedule for Children. *Journal of Clinical Child and Adolescent Psychology, 33*, 536-546.

Johnson, R. H., & Bond, G. L. (1950). Reading ease of commonly used tests. *Journal of Applied Psychology, 34*, 319-324. doi:10.1037/h0056258

Jones, K. D. (2010). The unstructured clinical interview. *Journal of Counseling & Development, 88*(2), 220-226.

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, T., Moreci, P., Williamson, D., & Ryan, N. (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child & adolescent Psychiatry, 36*(7), 980-988. doi:10.1097/00004583-199707000-00021

Kelleher, I., Connor, D., Clarke, M. C., Devlin, N., Harley, M., & Cannon, M. (2012). Prevalence of psychotic symptoms in childhood and adolescence: A systematic review and meta-analysis of population-based studies. *Psychological Medicine, 42*(9), 1857-63. doi:10.1017/S0033291711002960

Klinefelter, D., Pancoast, D. L., Archer, R. P., & Pruitt, D. L. (1990). Recent adolescent MMPI norms: T-score elevation comparisons to Marks and Briggs. *Journal of Personality Assessment, 54*, 379-389.

Klinge, V., Lachar, D., Grissell, J., & Berman, W. (1978). Effects of scoring norms on adolescent psychiatric drug users' and nonusers' MMPI profiles. *Adolescence, 13*(49), 1-11.

Klinge, V., & Strauss, M. E. (1976). Effects of scoring norms on adolescent psychiatric patients' MMPI profiles. *Journal of Personality Assessment, 40*, 13-17. doi:10.1207/s15327752jpa4001_3

Kovacs, M. (1985). The Children's Depression Study. *Psychopharmacol Bull, 21*, 995-998.

Marks, P. A., Seeman, W., & Haller, D. L. (1974). *The actuarial use of the MMPI with adolescents and adults*. Baltimore: Williams & Wilkins.

Mason, S. N., Bubany, S., & Butcher, J. N. (2012). Frequently asked questions: Gender differences on personality tests. Retrieved from www.umn.edu/mmpi

Merikangas, K. R., He, J. P., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., Benjet, C., Georgiades, K., & Swendsen, J. (2010). Lifetime prevalence of mental disorders in U.S. adolescents: Results from the National Comorbidity Survey Replication-Adolescent Supplement (NCS-A). *Journal of American Academy of Child and Adolescent Psychiatry, 49*(10), 980-989. doi:10.1016/j.jaac.2010.05.017

Miles, S., Fulbrook, P., & Mainwaring-Magi, D. (2018). Evaluation of standardized instruments for use in universal screening of very early school-age children: Suitability, technical adequacy, and usability. *Journal of Psychoeducational Assessment, 36*(2), 99-119. doi:10.1177/0734282916669246

Morey, L. C. (1991). *Personality Assessment Inventory: Professional manual*. Lutz: Psychological Assessment Resources.

Morey, L. C. (2007). *Personality Assessment Inventory: Professional manual* (2nd ed.). Lutz: Psychological Assessment Resources.

Moyle, W. (2001). Unstructured interviews: Challenges when participants have a major depressive illness. *Journal of Advanced Nursing, 39*(3), 266-273. doi:10.1046/j.1365-2648.2002.02273.x

Nock, M. K., Green, J. G., Hwang, I., McLaughlin, K. A., Sampson, N. A., Zaslavsky, A. M., & Kessler, R. C. (2013). Prevalence, correlates, and treatment of lifetime suicidal behavior among adolescents: Results from the National Comorbidity Survey Replication Adolescent Supplement. *JAMA Psychiatry, 70*(3), 300-310. doi:10.1001/2013.jamapsychiatry.55

North, C. S., Pollio, D. E., Thompson, S. J., Ricci, D. A., Smith, E. M., & Spitznagel, E. L. (1997). A comparison of clinical and structured interview diagnoses in a homeless mental health clinic. *Community Mental Health Journal, 33*, 531-543. doi:10.1023/A:1025052720325

Otto, R. K., Greenstein, J. J., Johnson, M. K., & Friedman, R. M. (1992). Prevalence of mental disorders among youth in the juvenile justice system. In J. J. Cocozza (Ed.), *Responding to the mental health needs of youth in the juvenile justice system*. Seattle: National Coalition for the Mentally Ill in the Criminal Justice System.

Odgers, C. L., Burnette, M. L., Chauhan, P., Moretti, M. M., & Reppucci, D. (2005). Misdiagnosing the problem: Mental health profiles of incarcerated juveniles. *The Canadian Child and Adolescent Psychiatry Review,* 14, 1-29.

Pancoast, D. L., & Archer, R. P. (1988). MMPI adolescent norms: Patterns and trends across four decades. *Journal of Personality Assessment, 52*, 691-706. doi:10.1207/s15327752jpa5204_9

Rettew, D. C., Doyle Lynch, A., Achenbach, T. M., Dumenci, L., & Ivanova, M. Y. (2009). Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *International Journal of Methods in Psychiatric Research, 18*(3), 169-184. doi:10.1002/mpr.289

Rogers, R. (1986). *Conducting insanity evaluations*. New York: Van Nostrand Reinhold.

Rogers, R. (2001). Nature of diagnostic and structured interviewing. In R. Rogers (Ed.), *Handbook of Diagnostic and Structured Interviewing* (p. 3-35). New York: Guilford Press.

Rogers, R. (2003). Standardizing DSM-IV diagnoses: The clinical applications of structured interviews. *Journal of Personality Assessment, 81*(3), 220-225. doi:10.1207/S15327752JPA8103_04

Rogers, R. (2008). Detection strategies for malingering and defensiveness. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., pp. 14-35). New York: Guilford Press.

Rogers, R., Sharf, A. J., Carter, R. M., Henry, S. L., Williams, M. M., & Robinson, E. V. (2015). Validity and representative data of the MRCI with legally involved juveniles. *Assessment*. doi: 10.1177/1073191115621792

Rogers, R., & Shuman, D. W. (2000). *Conducting insanity evaluations* (2nd ed.). New York: Guilford Press.

Rogers, R., Williams, M. M., Winningham, D. B., & Sharf, A. J. (in press). An Examination of PAI Clinical Descriptors and Correlates in an Outpatient Sample: Tailoring of Interpretive Statements. *Journal of Psychopathology and Behavioral Assessment*. doi:10.1007/s10862-017-9627-5

Rogers, R., & Wupperman, P. (2006). Diagnostic interviews. In S. Ayers, A. Baum, C. McManus, S. Newman, K. Wallston, J. Weinman & R. West (Eds.), Cambridge Handbook of Psychology, Health and Medicine (2nd Edition). Cambridge UK: Cambridge University Press.

Russell, M. A., & Marston, E. G. (2010). Profiles of mental disorder among incarcerated adolescent females. *Court Review: The Journal of the American Judges Association,* Paper 332. Accessed at: http://digitalcommons.unl.edu/ajacourtreview/332

Shedler, J. (2002). A new language for psychoanalytic diagnosis. *Journal of the American Psychoanalytic Association, 50,* 429-456. doi:10.1177/00030651020500022201

Sletten, I. W., Ulett, G., Altman, H., & Sundland, D. (1970). The Missouri standard system of psychiatry (SSOP): Computer generated diagnosis. *Archives of General Psychiatry, 23,* 73-79. doi: 10.1001/archpsyc.1970.01750010075014

Sharf, A. J., Rogers, R., Williams, M. M., & Drogin, E. Y. (in press). Evaluating juvenile detainees' Miranda misconceptions: The discriminant validity of the Juvenile Miranda Quiz. *Psychological Assessment*. doi:10.1037/pas0000373

Siefert, C. J., Sinclair, S. J., Kehl-Fie, K. A., & Blais, M. A. (2009). An item-level psychometric analysis of the Personality Assessment Inventory: Clinical scales in a psychiatric inpatient unit. *Assessment, 16*(4), 373-383. doi:10.1177/1073191109333756

Stein, S. J. (1987). Computer-assisted diagnosis in children's mental health. *Applied Psychology: An International Review, 36,* 343-355. doi:10.1111/j.1464-0597.1987.tb01196.x

Stokes, J. M., Pogge, D. L., & Archer, R. P. (2018). Comparisons between Minnesota Multiphasic Personality Inventory-Adolescent-Restructured Form (MMPI-A-RF) and MMPI-A in adolescent psychiatric inpatients. *Psychological Assessment, 30*(3), 370-382. doi:10.1037/pas0000488

Summerfeldt, L. J., & Antony, M. M. (2002). Structured and semi-structured diagnostic interviews. In A. M. Antony (Ed.), *Handbook of assessment and treatment planning for psychological disorders* (p. 3-37). New York, NY: Guilford Press.

Tarolla, S. M., Wagner, E. F., Rabinowitz, J., & Tubman, J. G. (2002). Understanding and treating juvenile offenders: A review of current knowledge and future directions. *Aggression and Violent Behavior, 7*, 125-143. doi:10.1016/S1359-1789(00)00041-0

Tellegen, A., & Ben-Porath, Y. S. (2008, 2011). *Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF): Technical manual*. Minneapolis: MN: University of Minnesota Press.

Thatte, S., Makinen, J. A., Nguyen, H. N. T., Hill, E. M., & Flament, M. F. (2013). Partial hospitalization for youth with psychiatric disorders: Treatment outcomes and 3-month follow-up. *Journal of Nervous and Mental Disease, 201*(5), 429-434. doi:10.1097/NMD.0b013e31828e1141

Turner, S. M., Hersen, M., & Heiser, N. (2003). The interviewing process. In M. Hersen & S. M. Turner (Eds.), *Diagnostic interviewing* (3rd. ed., p. 3-20). New York, NY: Kluwer Academic/Plenum.

Veltri, C. O. C., Graham, J. R., Sellbom, M., Ben-Porath, Y. S., Forbey, J. D., O'Connell, C., … White, R. S. (2009). Correlates of MMPI-A Scales in acute psychiatric and forensic samples. *Journal of Personality Assessment, 91*(3), 288-300. doi:10.1080/00223890902794374

Ward, C. H., Beck, A. T., Mendelson, M., Mock, J. E., & Erbaugh, J. K. (1962). The psychiatric nomenclature: Reasons for diagnostic disagreement. *Archives of General Psychiatry, 7*(3), 198-205. doi:10.1001/archpsyc.1962.01720030044006

Ward, L. C., & Ward, J. W. (1980). MMPI readability reconsidered. *Journal of Personality Assessment, 44*, 387-389.

Weiner, Z., Reich, W., Herjanic, B., Jung, K. G., Amado, H. (1987). Reliability, validity, and parent-child agreement studies of the Diagnostic Interview for Children and Adolescents (DICA). *Journal of the American Academy of Child and Adolescent Psychiatry, 26*, 649-653.

Westen, D., & Shedler, J. (1999). Revising and assessing Axis II, Part I: Developing a clinically and empirically valid assessment method. *American Journal of Psychiatry, 156*, 258-272.

Wilkinson, G. S., & Robertson, G. J. (2006). *Wide Range Achievement Test – Fourth*

Edition (WRAT-4) professional manual. Lutz, FL: Psychological Assessment Resources.

Williams, C. L. (1986). MMPI profiles from adolescents: Interpretive strategies and treatment considerations. *Journal of Child and Adolescent Psychology, 3*, 179-193.

Williams, C. L., & Butcher, J. N. (1989a). An MMPI study of adolescents: I. Empirical validity of the standard scales. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1*, 251–259.

Wrobel, N. H., & Lachar, D. (1992). Refining adolescent MMPI interpretations: Moderating effects of gender in prediction of descriptions from parents. *Psychological Assessment, 4*(3), 375-381. doi:10.1037/1040-3590.4.3.375

Zeanah, C. H., & Myint, M. T. (2017). Editorial: Minding the gap - research on sexual minority and gender nonconforming children and adolescents. *The Journal of Child Psychology and Psychiatry, 58*(11), 1177-1179. doi:10.1111/jcpp.12836

Zimmerman, M. (2002). *Psychiatric diagnostic screening questionnaire manual*. Lose Angeles: Western Psychological Services.

Zimmerman, M., & Mattia, J. L. (1999). The reliability and validity of a screening questionnaire fro 13 DSM-IV Axis I disorders (the psychiatric diagnostic screening questionnaire) in psychiatric outpatients. *Journals of Clinical Psychiatry, 60*, 677-683. doi:10.4088/JCP.v60n1006

Zimmerman, M., & Mattia, J. I. (2001). The Psychiatric Diagnostic Screening Questionnaire: Development, reliability and validity. *Comprehensive Psychiatry, 42*, 175-189. doi:10.1053/comp.2001.23126