

EXAMINING THE APPROPRIATENESS OF USING NON-AGE STRATIFIED NORMS ON
THE MMPI-A-RF

A THESIS

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF ARTS

BY

YASMINE GRIFFIN

DR. TAYLA T.C. LEE - ADVISOR

BALL STATE UNIVERSITY

MUNCIE, INDIANA

JULY 2023

ABSTRACT

THESIS PROJECT: Examining the Appropriateness of Using Non-Age Stratified Norms on the MMPI-A-RF

STUDENT: Yasmine A. Griffin

DEGREE: Master of Arts

COLLEGE: Sciences and Humanities

DATE: JULY 2023

PAGES: 105

The present study sought to investigate whether the Minnesota Multiphasic Personality Inventory-Adolescent-Restructured Form (MMPI-A-RF; Archer et al., 2016) should use age stratified norms. The MMPI-A-RF is one of the most commonly used assessments for adolescents, but it does not use age stratified norms or interpretations. However, there is evidence to suggest that adolescence is a period of rapid and immense biological, emotional, cognitive, and social change, which could indicate the need for age-stratified norms or interpretations. I examined this possibility using Multiple Indicators Multiple Causes (MIMIC) models for each of the Restructured Clinical (RC) scales to garner an understanding of whether age significantly influenced endorsement of specific items (i.e., differential item functioning; DIF) or the latent variable representing the scale score. I also conducted tests of moderation in a regression context to assess whether age moderated the relations between RC scale scores and selected outcome measures. Results of the MIMIC analyses suggested that while there were several items with DIF, these items did not result in age being related to the latent variable representing the scales' scores. The one exception was RCd, where there were 2 items with DIF (Items 8 [Hard to keep mind on task; $\beta = .17, p < .05$] and 190 [Happy most of the time; $\beta = .12, p < .05$]), and age demonstrated a consistent association with the latent variable representing RCd ($\beta = -.16$ to $-.14$). Results of the tests of moderation suggested that age did not significantly

interact with scores on the RC scales to impact scores or symptom counts on the selected criterion measures. Overall, results of the current analyses do not support the need for age stratified norms or interpretations for the RC scales, with the exception of RCd. Items and scores on RCd may have demonstrated being influenced with age because younger adolescents experience greater mood fluctuation than older adolescents (Klimstra et al., 2015), and as a result, RCd may be sensitive to this heightened emotional arousal in younger adolescents. If additional research supports an influence of age on RCd, age stratified norms that compare adolescent reports to reports from other similarly aged individuals or age-graded interpretations that consider age related vulnerabilities towards emotional arousal, should likely be adopted for routine use with the instrument. However, given that this study is the first of its kind for the MMPI-A-RF RC scales, results should be viewed cautiously and within the context of the limitations of the current study.

Examining the Appropriateness of Using Non-Age Stratified Norms on the MMPI-A-RF

The Minnesota Multiphasic Personality Inventory-Adolescent-Restructured Form (MMPI-A-RF; Archer et al., 2016) is a commonly used instrument assessing the psychological and emotional well-being of adolescents 14- to 18- years old. However, the MMPI-A-RF differs from other psychological measures used with adolescents, as it does not use age stratified norms. While there is no available evidence to support the need for age stratified norms, research has shown that adolescence is a period of rapid and immense development, which could impact the accuracy of interpretations made based on the measure's scales. Because of this possibility, the goal of the current study was to investigate whether age influences item responses or scale scores, as well as whether scores predict outcomes differentially based on age. Evidence supporting the presence or absence of age differences would provide evidence indicating whether the MMPI-A-RF should use age stratified norms.

Adolescent Development

Adolescence is a developmental period characterized by biological, cognitive, and social changes that are critical developmental transitions between child- and adulthood (Steinberg, 2005; Sanders, 2013). Adolescence begins with changes in biology and ends in social transitions into adulthood marked by sociological terms such as marriage, education completion, and entering the work force (Smetana et al., 2006).

Biologically, adolescence is a departure from the childlike stature and physique associated with childhood. The biological transition from childhood into adolescence is commonly associated with pubertal maturation, which is one of the most impactful changes. Pubertal maturation is defined by distinct physical growth (e.g., height, weight, increase hormone production, etc.) and sexual development (e.g., growth of pubic hair, maturation of the

genitals, first menstruation) (Berenbaum et al., 2015; Christie & Viner, 2005; Reiter & Root, 1975;). Puberty has been found to have an influence on an adolescents' social and affective processing (Crone & Dahl, 2012). For example, increases in gonadal hormones (e.g., testosterone, estradiol) can sensitize the brain's reward system, making adolescents more reward oriented and more likely to engage in risk-taking (Peper & Dahl, 2013). Puberty can also impact an adolescent's social standing with their peers, interactions with their family members, ideas of self, and psychological well-being, which is reflected in the other domains of adolescent development (i.e., cognitive, social, and emotional development) (Morris et al., 2007; Steinberg, 2005; Steinberg & Morris, 2001).

Other biological changes that occur during adolescence center around the brain, and these brain developments facilitate the cognitive features characteristic of adolescence. During adolescence, there is an increase in the links in the entire brain and increased connectivity through myelination of the nerve fibers and localized synaptic pruning in the frontal lobe (Arian et al., 2022; Steinberg, 2005). These changes in the brain give rise to cognitive features that allow for increasingly complex thought processes (e.g., abstract thinking, hypothesizing) and the development of more complex cognitive functions (e.g., executive functioning, decision-making) (Arian et al., 2022; Steinberg, 2005). This is a stark difference from childhood, which is characterized by more concrete thinking (i.e., organizational and logical thinking about concrete events) (Steinberg, 2005).

In adolescence, biological changes in the body are also stimulating affective behaviors and risk-taking, but the brain and other regulatory systems have not yet fully developed to regulate and moderate emotionality (Arian et al., 2022; Steinberg, 2008). Specifically, there are immature connections between the limbic system, prefrontal cortex, and amygdala compared to

adulthood, a period in which connectivity is done developing. This is reflected in the fact that while by the age of 15 there is not a difference between adolescents' and adults' decision-making, the addition of complex feelings makes it more difficult for adolescents to think through the costs of their decisions (Arian et al., 2022). As a result of this heightened emotional responsivity in adolescence, there is a need for emotional developments specifically in emotion regulation—a set of internal and external processes to manage emotions (Morris et al., 2007). Emotion regulation differs from childhood into adolescence in that in childhood, emotion management is primarily reliant on caregivers to help with regulating one's emotions and intervening when they are unable to manage (Morris et al., 2007). In adolescence, a youth's emotion regulation begins to be influenced by other social agents, such as their peers (Morris et al., 2007).

These changes in emotion regulation can also partly be attributed to the social developments that occur during adolescence. During childhood, parental and authority figures help guide decision making and identity development; however, in adolescence, there is a departure from relying on these individuals and a shift to relying on peers. This reliance on peer relationships becomes increasingly important, with young adolescents being the most strongly influenced by their similar aged counterparts (Knoll et al., 2015). This sensitivity towards peers puts adolescents at a greater susceptibility to either be positively or negatively influenced by peers, which can be seen in the fact that research has suggested that adolescents' evaluations of situational risk are influenced by peers (Knoll et al., 2015). Also, social influences can impact academic achievement, engagement in prosocial or problematic behaviors (e.g., alcohol use, delinquency) and psychological well-being (Steinberg & Morris, 2001). Overall, this shift in

influence paired with the changes in emotion regulation and decision making is associated with greater conflicts between adolescents and their families (Steinberg & Morris, 2001).

Stages of Adolescent Development

The changes that occur in adolescence do not happen all at once; rather, they occur throughout adolescence with specific challenges and growths. Indeed, there is evidence to support that adolescence is comprised of three developmental stages: early, middle, and late adolescence (Christie & Viner, 2005; Smetana et al., 2006). Early adolescence, which occurs approximately between the ages of 10 and 13, is characterized by the beginning of puberty, as well as increased emotional arousal, sensation-seeking, and reward orientation in comparison to children (Steinberg, 2005). It is during this stage of development, that one sees improvements in deductive reasoning, information processing, and abstract thinking in comparison to children (Steinberg, 2005; Steinberg & Morris, 2001), who experience more positive emotions and whose behaviors are more controlled by their parents (Denham, 2019). However, while there are improvements in reasoning and thinking processes, early adolescents experience difficulties in decision making in comparison to older adolescents, as early adolescents' decision making is primarily influenced by high arousal, increased peer influence, and lower impulse control in comparison to older adolescents and emerging adults (Crone & Dahl, 2012; Steinberg, 2005). Thus, adolescents in this stage are at risk for sensation-seeking, risk-taking, and generally reckless behaviors (Steinberg, 2005).

Middle adolescence, which occurs approximately between the ages of 14 and 17, is when self-discovery begins (Steinberg & Morris, 2001). Adolescents are beginning to attempt to describe themselves, but the descriptions may be occasionally discrepant ("e.g. shy with friends"; Backes et al., 2019; Steinberg & Morris, 2001). The challenges found in early

adolescence carry into middle adolescence in that middle adolescents have a lower threshold for regulating affect and behavior than those in later stages of adolescence (Steinberg, 2005). In middle adolescence, the adolescent also becomes more self-involved and begins to have concerns about how they appear to others, especially their peers (Christie & Viner, 2005; Knoll et al., 2015; Steinberg & Morris, 2001). These changes frequently lead to increased conflict with caregivers, though these tend to gradually decrease towards the end of middle adolescence.

Late adolescence, which occurs from 17 to 19 years old, is much closer to what would be considered emerging adulthood because the brain has continued to develop, especially in the frontal lobe, and their emotions has begun to stabilize in comparison to early and middle adolescents who experience intense emotions without the biological and cognitive means of regulating them (Steinberg, 2005). This implies younger adolescent experience greater mood fluctuation compared to older adolescents who are liable to be able to delay gratification and think through decisions (Klimstra et al., 2015; Steinberg, 2005). Older adolescents also have increased abilities for expressing themselves. Their identity is becoming more concrete, and they begin to have greater concerns about their future. In comparison to middle adolescence, late adolescent identity development is not as heavily influenced by peers and is more complex, abstract, differentiated, and organized (Christie & Viner, 2005; Knoll et al., 2015; Steinberg & Morris, 2001). Identity is also based more on one's own personal beliefs, and less on social comparisons or their parents (Sanders, 2013; Steinberg & Morris, 2001). Alongside this, adolescents in this stage have a greater capacity for perspective taking and acting independently than their younger peers (Dumontheil et al., 2010; Steinberg, 2005), as they have established more responsible behaviors and matured personal values (Sanders, 2013). However, their capabilities can be differentiated from that of adults, as their brains are still developing. They

also lack specific life experiences characteristic of adulthood, and their decision making is still in-part influenced by peers and emotional factors.

Psychological Assessment of Adolescents

From the previous section, it should have been clear that adolescence is a period of self-discovery, new influences, and general susceptibility to socioemotional arousal. While developmentally expected and appropriate, the nature and rapidity of these changes poses a challenge to personality and psychological assessment of adolescents. The reason being is because these changes can make scores on measurements of adolescents' characteristics and abilities susceptible to change across time (Archer, 2016). Hence, responses that may be true one day may not also be true the next time an adolescent is given the same measure. This change in scores on a measure due to development is referred to as maturation effects. Maturation effects, a process where an individual varies across time due to biological or psychological processes within the individual, negatively impacts the ability of assessments to accurately predict psychological well-being past the time in which the assessment was given (Archer, 2016). This susceptibility towards maturation effects limits the utility of certain tests that lack sensitivity for capturing the transient organization of the personality during adolescence. This poses a problem for assessment because if the measure is not sensitive to the developmental changes that can occur during adolescence, erroneous assumptions can be made based on what could be a developmentally appropriate response. In other words, inaccurate interpretations can be made that would impact important decisions about the care and treatment of an adolescent.

One way of accounting for change in scores across developmental periods or ages is to base standardized scores and interpretations on stratified normative samples. Essentially, normative data or norms serve the role of telling the assessor where the individual stands relative

to a comparison group and allows for direct comparisons between the individual's performance and that comparison group's performance (Anastasi & Urbina, 1997; Committee on Psychological Testing et al., 2015). There are many types of norms; however, the norms that will be of focus in this paper are stratified norms. Norms as a whole are supposed to be representative of the population in which the test or measure is used (Committee on Psychological Testing et al., 2015). Stratified samples or stratified norms are when the norm sample is broken down into groups with particular demographic characteristics represented in the population (e.g., age, race, parental education, socioeconomic status, or geographical region of the country; Committee on Psychological Testing et al., 2015). These stratifications allow for a better approximation of these features in proportion to the population and comparisons within a group. Within-group comparisons or norms involve evaluating the individual's performance in relation to the most comparable standardization group (Anastasi & Urbina, 1997). For example, comparing a child's score with children of the same age or in the same grade.

Cognitive and behavioral assessments of children and adolescents typically use age and or gender stratified norms. For example, the Weschler Intelligence Scale for Children-Fifth Edition (WISC-V; Weschler, 2014) is a measure of intellectual ability and cognitive domains that impact performance. The WISC-V uses stratified norms with age groups of children from 6 to 16 years old to allow for comparisons between same-age children because cognitive ability develops rapidly in childhood and adolescence. The WISC-V used stratified sampling to ensure that age groups were representative of the October 2012 U.S. Census Bureau data "with respect to age, sex, race/ethnicity, parent education level, and geographic region" (Benson & Keith, 2017). Using stratified norms, such as the one's used in the WISC-V, are intended to allow for

fair interpretations, as each individual's performance would be compared to similarly aged individuals who are hypothesized to have similar cognitive skills.

As another example, the Behavioral Assessment System for Children - Self Report of Personality (BASC-SRP; Reynolds & Kamphaus, 2015) is an assessment tool that can be used for both children (SRP-C) ages 8 to 11 and adolescents (SRP-A) ages 12 to 21 to assess their behavioral, psychological, and emotional wellbeing. The SRP is a component of the BASC-3 suite of instruments, which measure a youth's perceptions of and feelings toward parents, peers, school, and their own behavior. The SRP uses age and gender stratified norms. The reason the developers of the BASC use gender and age stratified norms is because analyses conducted during the development of the BASC-2 showed age and gender differences on a number of scales. Specifically, significant differences between age groups were found on the Attitude to Teachers, Atypicality, Hyperactivity, Interpersonal Relations, Locus of Control, and Self-Reliance scales.¹ To account for differences on these scales, the test's authors determined that it would be appropriate to use the age groupings from the BASC for the BASC-2 Norms, and those were carried over onto the BASC-3.

The Minnesota Multiphasic Personality Inventory (MMPI) Family

The WISC and the BASC are among the most widely used assessment instruments used with adolescents; however, other widely used assessment tools with this age group are the Minnesota Multiphasic Personality Inventory (MMPI) family of instruments (Archer et al., 1991; Archer & Newsom, 2000; Cashel, 2002). The MMPI family of instruments are meant to assess an individual's psychological and emotional well-being. The first member of the family is the original MMPI, which was primarily intended for use with adults (Archer, 1984). However, it

¹ Specific statistics such as effect sizes were not included in the manual.

became a popular instrument for assessing adolescents; specifically, this early version of the MMPI was used with “bright children” as young as 12 years old (Archer, 1984). However, there were some concerns about the use of the MMPI with children and adolescents, as it was developed with and normed on adults. As a result, many of the items on the MMPI were irrelevant or inappropriate for adolescents (Graham et al., 2022). Furthermore, several studies discovered the adult MMPI norms over pathologized adolescent test-takers (Archer, 1984; Marks et al., 1974). Although attempts were made to add adolescent norms to the MMPI, these norms tended to under pathologize adolescents experiencing psychological difficulties (Archer et al., 1986; Klinge et al., 1978). Finally, there were concerns regarding the interpretative statements derived from scores for adolescents, as there was confusion about whether to use the adult interpretative statements or the descriptors developed specifically for adolescents (Archer, 1984).

The MMPI was updated in 1989 for adults (Butcher et al., 1989). However, because of the need for representative adolescent norms and adolescent-specific scales (e.g., a scale measuring school functioning), the committee responsible for developing the MMPI-2 decided to develop a form of the MMPI specifically for use with adolescents (Archer, 2016; Butcher & Pope, 1992). This effort resulted in the Minnesota Multiphasic Personality Inventory - Adolescent (MMPI-A; Butcher et al., 1992). Like its predecessor, the MMPI-A was meant to assess an individual’s psychological and emotional well-being, but incorporated more appropriate and relevant items for an adolescent population, and interpretations were based on adolescent norms.

Archer and Newsome (2000) found that the MMPI-A became one of the most common assessment tools used with adolescents. However, despite its benefits, there were a few limitations. These derived primarily from the fact that the developers of the MMPI-A wanted to

maintain continuity of the Validity and Clinical scales of the original MMPI given their large research base (Archer, 2016). However, as a result, the MMPI-A maintained limitations found for the MMPI, including issues with scale multidimensionality, content heterogeneity, and item overlap (Archer et al., 2016). These limitations led to extremely large intercorrelations between many of the MMPI-A scales, which, in turn, led to problems with the discriminant validity of the scales' scores (Archer et al., 2016). Alongside the previously mentioned limitations, the MMPI-A was also relatively long with 478 items, which posed a challenge to the attention span and concentration abilities of many of the adolescent test-takers (Archer et al., 2016).

The MMPI-A-RF, a restructured version of the MMPI-A published in 2016, was developed to address the limitations of the MMPI-A (Archer et al., 2016). To address concerns about the length of the test, the MMPI-A-RF was reduced to 241 items. To account for the psychometric problems of substantive scales, Archer and colleagues (2016) used a restructuring process similar to that used for the adult version of the instrument. Namely, when addressing these types of concerns for the MMPI-2, the process began with Tellegen and colleagues (2003) developing nine Restructured Clinical (RC) scales. These scales were intended to retain the strengths of the original MMPI Clinical scales while also reflecting clinically meaningful variance of the MMPI-2 item pool in a more psychometrically appropriate manner.

Subsequently, additional scales were developed using similar methods for the MMPI-2, and these efforts resulted in a restructured version of the MMPI-2 intended for use with adults, the Minnesota Multiphasic Personality Inventory - 2 - Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2011; Tellegen & Ben-Porath, 2011). Research suggested that the MMPI-2-RF scales were successful in addressing psychometric issues observed for the MMPI-2 (Sellbom, 2019).

The end result of the restructuring process for the MMPI-A was an instrument containing 48 scales in total, including six Validity scales assessing response styles and 42 scales assessing substantive constructs related personality and the experience of psychopathology (Graham et al., 2022). The substantive scales are arranged in a hierarchical manner and include the Higher-Order (HO), Restructured Clinical (RC), Specific Problems (SP), and Personality Psychopathology Five (PSY-5) scales (Archer et al., 2016). The H-O scales represent the broadest level of the hierarchy and assess major dimensions of psychological functioning, including internalizing, externalizing and thought dysfunctions. The nine RC scale assess mid-level aspects of psychopathology experienced by adolescents and were meant to reflect the same constructs assessed on adult versions of the RC scales. Given these scales represent the key aspects of psychopathology assessed by the MMPI-A-RF, they will be the focus of this study and will be described in more detail below. At the narrowest level of measurement on the MMPI-A-RF are the 25 SP scales that represent the clinically important facets of the RC scales, or clinically significant factors not assessed by the RC scales. Lastly, adjacent to the MMPI-A-RF hierarchy are the Personality Psychopathology-Five (PSY-5) scales, which are meant to reflect the constructs in Harkness and McNulty's (1994) model of maladaptive personality traits.

Initial evidence presented in the *MMPI-A-RF Manual for Administration, Scoring, Interpretation, and Technical Manual* (Archer et al., 2016), as well as in peer-reviewed research (Sharf & Rogers, 2020; Stokes et al., 2018), is supportive of the reliability and validity of MMPI-A-RF scale scores. Archer and colleagues (2016) provided test-retest and internal consistency (α) coefficients for all MMPI-A-RF substantive scales, though I will discuss only the RC scales given they are the focus of the current study. For the RC scales, test-retest reliabilities ranged from .56 (RC8) to .82 (RCd), with α coefficients ranging from .45 to .83 considering minor

differences between boys and girls. However, it should be noted that lower test-retest coefficients and lower α coefficients could be the result of range restriction in the normative sample and relatively small number of items for some of the scales impacting the stability coefficients (Graham et al., 2022).

With regards to validity, Archer et al. (2016) provided intercorrelation data that suggest that the RC scales are measuring similar characteristics to those assessed in the MMPI-A Clinical and Content Scales, which provides support for the construct validity of the MMPI-A-RF scales. For example, RC1, a scale measuring somatic complaints, demonstrated associations with scores on Clinical Scale 1 and the Health Concerns Content scale of the MMPI-A, both of which also measure somatic difficulties (Archer et al., 2016; Graham et al., 2022). The correlates provided by Archer et al. (2016) also provide support for the convergent and discriminant validity of the RC scales. For example, RCd, a scale assessing generalized distress and demoralization, significantly and positively related with emotional distress factors such as sadness, depression, and anxiousness, which aligns with the idea that demoralization refers to one's inability to cope and feelings of helplessness and hopelessness. (Archer et al., 2016). There are some exceptions to the intercorrelations providing support for the convergent and divergent validity of the MMPI-A-RF. For example, Archer and colleagues (2016) were not able to correlate RC3, a measure of cynical world views, with other measures of cynical attitudes; as a result, more research is needed to establish RC3 as a measure of cynicism (Graham et al., 2022).

Outside of the MMPI-A-RF Manual, there is growing research supporting the MMPI-A-RF scale scores' validity, especially the RC scales. For example, Stokes et al. (2018) found that the RC scales were significantly less intercorrelated and more independent than the MMPI-A Clinical scales, suggesting the MMPI-A-RF adequately addressed the issues of high

intercorrelations between the MMPI-A Clinical scales. Another example comes from Sharf and Rogers (2020), who found evidence to support the convergent and discriminant validity of the MMPI-A-RF RC Scales. Specifically, these authors found several the RC scales were correlated with specific disorders relevant to what the scale measured. For example, RC4, a scale measuring antisocial behavior, correlated with a range of externalizing disorders, demonstrating good convergent validity. With regards to discriminant validity, many of the RC scales did not have significant associations with disorders unrelated to the construct they were intended to measure. For example, RC7, a measure of negative emotionality, significantly correlated with Major Depressive Disorder and Generalized Anxiety Disorder, but symptoms of these disorders were not strongly associated with RC4, a measure of antisociality.

Despite this initial positive support for the MMPI-A-RF, one potential limitation of the instrument is that it uses the same norms as the MMPI-A. The normative sample of the MMPI-A-RF is a subset of the MMPI-A normative sample because test's authors concluded that the MMPI-A norms collected in the late 1980s and early 1990s were similar to the more recent adolescent samples (Archer et al., 2016). This suggested that the use of the norms in the MMPI-A-RF was appropriate and should yield profiles appropriate for the current samples of adolescents. The MMPI-A-RF's developers also retained the use of non-stratified norms, as was done for the MMPI-A (Archer et al., 2016). Although the data were not published, Graham et al. (2022) reported that during the MMPI-A's development, data analyses indicated that there were only minor differences in scale scores between age groups within the MMPI-A normative sample, suggesting that stratified norms were not needed. However, there may be some problematic assumptions with making this determination based on the absence of mean score differences.

Measurement Bias

To date, no research has examined the appropriateness of using non-stratified norms on the MMPI-A-RF. This is an important area of inquiry, however, given the MMPI-A-RF is used in various settings and high-stakes situations where it is imperative that interpretations of the scales are accurate and well-informed. As previously mentioned, not using stratified norms when assessing adolescents who are in a rapid period of development poses a risk of over- or under-pathologizing a particular group. In the psychological assessment literature, this is referred to as bias (Reynolds & Suzuki, 2012). Specifically, bias in psychological assessment occurs when a test systematically under- or over-represents a specific group's aptitude, ability, or standing on a characteristic (Reynolds & Suzuki, 2012). This is not to be misconstrued with fairness, which refers to the moral, philosophical, or legal issues people can disagree on (Reynolds & Suzuki, 2012). Bias is a statistically estimated quantity based on evidence, not opinion. Clinicians are ethically obligated to not only use fair assessment tools, but also unbiased assessment instruments (American Psychological Association, 2020).

As the data analyses were not made publicly available regarding why the most recent adolescent version of the MMPI does not use age stratified norms, it is not entirely clear what evidence supported the decision to not use age stratified norms. However, based on the reasoning for why one would use age stratified norms, it could be assumed that non-age stratified norms were used because there was a lack of age-group mean differences on scale scores. This may be a problematic decision because analyzing group means is not sufficient to determine whether or not a test is biased because of a limitations of mean difference testing (Reynolds & Suzuki, 2012).

The first limitation of using mean differences by themselves to suggest bias is that these tests actually only indicate that two groups are different with relation to performance (Reynolds & Suzuki, 2012). However, it may be that mean differences between groups captures some bias, but that it also includes other confounding variables, which would impact scores (Reynolds & Suzuki, 2012). As a result, one could not be sure of the amount of bias present or which group is favored (Reynolds & Suzuki, 2012). Furthermore, efforts to modify scoring or alter interpretation based on mean differences could create more bias as one does not know the reason for the differences.

Other limitations with mean difference tests are that they do not provide important statistical information about group differences and may not represent the entire population. Specifically, mean difference testing does not provide information regarding skewness, kurtosis, and variation (Reynolds & Suzuki, 2012). A norming sample can produce symmetric or normally distributed scores but also produce asymmetric or skewed scores for a particular group (Reynolds & Suzuki, 2012). Skew can tell whether the scores are clumping higher or lower on the scale and whether the mean is being pulled towards the clump of scores (Reynolds & Suzuki, 2012). Because mean difference testing does not provide information regarding symmetry, if asymmetry is present within a particular group's score distribution, results of mean difference tests could produce inaccurate means and, therefore, inaccurate mean differences (Reynolds & Suzuki, 2012). Kurtosis, on the other hand, refers to dispersion of scores relative to the peak of the distribution, which would not necessarily affect the mean but could misrepresent score dispersion (Reynolds et al., 2009). Specifically, two groups may have the same mean but vastly different standard deviations, which would suggest that one group may have greater variability

than the other. Because of this, simply testing for mean differences would not precisely inform the investigator of whether the groups actually differ.

There are alternative methods to difference tests for identifying whether there is a need to use stratified norms, and that is through examining measurement invariance and tests of moderation. Measurement invariance (MI) is a test of construct validity, which is whether a test measures the theoretical construct or trait it is intended to assess as predicted across groups (Reynolds & Suzuki, 2012).² MI requires that items on a test and the latent traits assessed by these items should not depend on group measurement or measurement factors (Van De Schoot et al., 2015). With that being said, MI would be able to inform whether groups are performing differently at the item level. If invariance across groups is present, it would suggest that an item or items are not as representative of the latent trait being assessed for one group versus another.

The second alternative method, tests of moderation, pertains to the errors in predictions made for people in different groups. Sometimes called tests of predictive validity, tests of moderation are used to inform whether different predictions can be made for one group versus another. If differences are found, it would suggest that being a member of a particular group comes with some degree of systematic error, or an unaccounted variable that is only observed in one of the groups. There are two main aspects to consider in tests of predictive validity - slope and intercept bias. Differences in either slope or intercept indicate that a regression equation predicts an outcome inaccurately for one group compared to another. Slope bias would indicate there is greater criterion validity for one group in comparison to another; whereas intercept bias

² The language of MI and tests of moderation often assumes the influential variable is categorical. I have adopted that language here for the sake of simplicity. However, MI and tests of moderation can test for potential influential variables that are continuous in nature, such as age. For continuous variables, the differences would occur across levels of the continuous variable, rather than across groups.

refers a test consistently over or under predicting criterion levels for one group in comparison to another.

Overall, these methods are better than mean difference testing for identifying bias because they provide information regarding why and how the scores function differently across groups. Both MI and tests of predictive validity provide the basic statistical information regarding skew, kurtosis, and variance. However, individually, MI provides information about whether the items of the measure are the cause of possible differences across groups (Schmitt & Kuljanin, 2008). Tests of differential validity provide a point of understanding where the predictive inaccuracies are occurring through assessing how scores on the measures relate with criterion variables across groups (Reynolds & Suzuki, 2012). Using these methods will allow me to assess whether or not the MMPI-A-RF has items or scale scores that function differently across ages, and, thus, to examine whether or not the instrument should use age stratified norms.

Current Study

As reviewed, adolescence is characterized by a number of developmental changes that pose a problem to accurate psychological assessment. To address this problem, many assessment measures use age-stratified norms to ensure that the interpretations that are being made are both accurate and developmentally appropriate. The MMPI-A-RF, a commonly used measure with adolescents, does not use age-stratified norms, and the appropriateness of this practice is understudied. Thus, the current study sought to investigate whether there is a need for age-stratified norms for the MMPI-A-RF using MI and tests of moderation to examine whether there are items and scale scores that function differently across ages. Investigating this problem is imperative as the MMPI-A-RF is used in both clinical and forensic settings where information

provided by the instrument's scales may influence important decisions regarding the care and treatment of an adolescent.

While there are 42 substantive scales on the MMPI-A-RF, the current study will focus on the nine Restructured Clinical (RC) scales, which are listed and described in Table 1. This is because these scales represent the core of the MMPI-A-RF and assess key aspects of psychopathology experienced by adolescents (Archer et al., 2016). I first used MI testing to investigate whether responses to items comprising the RC scales differed across age. These tests allowed me to assess whether responses to a specific item on a scale are dependent on levels of age. Evidence of there being many invariant items would suggest the use of non-age stratified norms is inappropriate. These analyses also allowed me to examine whether age was consistently related to a latent representation of the scale's scores. Such associations would also indicate the need for age stratified norms (or age stratified interpretations). Next, I used tests of moderation to assess whether RC scale scores prediction of conceptually relevant criterion measures was influenced by age. Outcomes identified for use in this study for each RC scale are listed in Table 1. These tests would indicate whether there is greater criterion validity for one age over another (i.e., slope differences) and whether there is consistent over- or under-prediction of criterion for one age over another (i.e., intercept differences).

I hypothesized some item level invariance would be found with relation to age. I also hypothesized that there would be evidence to support predictive invalidity with relation to age. This is because, as previously discussed, adolescence is a period of great developmental change. An individual in the early or middle stages of adolescence would be different from an individual in the later stages of adolescence, as the individual in the later stages of adolescence has experienced specific growths in their biological, cognitive, and emotional functioning. These

rapid developments impact the assessment and evaluation of personality in adolescence, as what an adolescent endorses during one stage of adolescence may not be endorsed by an adolescent in a different stage of adolescence. For example, those in early adolescence may endorse greater emotional arousal and lower attentional capacities than someone in later adolescence. This difference in endorsement does not mean that the early adolescent has greater psychological dysfunction in comparison to the older adolescent, as developmentally someone in early adolescence is predisposed to higher levels of emotional dysregulation and inattentiveness when compared to older adolescents who have experienced biological, cognitive, and emotional changes where these abilities are developed. In this example, using a test without age-stratified norms would over-pathologize the experiences of the younger adolescent. While I believed that I would find some age differences either through tests of MI or tests of moderation, my hypotheses for both tests were non-specific because research regarding age differences and the MMPI-A-RF is limited and, as a result, unable to inform specific hypotheses.

Methods

Participants and Procedure

The current study used de-identified data collected from a juvenile residential facility located in the Midwest United States. Upon entering the program, residents at this facility underwent a standardized diagnostic evaluation where the residents completed a number of assessment measures and interviews. Data from the resident's medical records, including the responses to the assessment measures included in the diagnostic evaluation, were coded by the research team after the evaluation was completed. Data from residents who completed the MMPI-A-RF and other measures identified for use in this study (Table 1) were used in the current study. The original data set included 391 adolescents with ages ranging from 13 to 18

years of age ($M = 15.27$, $SD = 1.18$). The sample was primarily male ($n = 230$, 58.8%), with 40.9% of the sample being female. Residents were primarily White ($n = 298$, 76.2%), with 10% of the sample identifying as Black, 9.2% identifying as biracial, 3.6% identifying as Hispanic, and 1% identifying as Indigenous/Native.

To reduce error variance in analyses, cases were excluded from analyses if they had an invalid MMPI-A-RF. Archer and colleagues (2016) defined invalid profiles as those that exceed recommended cut-scores on the MMPI-A-RF Validity scales, including scores on Variable Response Inconsistency- $r > 75T$, True Response Inconsistency- $r > 75T$, Combined Response Inconsistency > 75 , or Infrequent Responses $> 80T$. Following the recommendations of Reynolds and Kamphaus (2015), participants were also excluded if they present invalid BASC-3 profiles (i.e., omitting responses on three or more items for a scale, V-index indicating extreme caution, F-index indicating extreme caution, or L-index indicating extreme caution). After excluding participants, 58 residents were removed from the sample, leaving a sample of 333 residents between the ages of 13 and 18 years of age ($M = 15.29$, $SD = 1.18$). As expected, the sample was primarily male ($n = 199$) and White ($n = 252$, 75.7%). Demographics between the original sample and the sample with valid profiles were comparable as trends for mean age, sex, and race were upheld.

Although the valid sample came to include profiles for 333 residents, this number does not reflect the individuals that only had valid BASC and valid MMPI profiles nor the number of individuals who had specific MINI-KID modules. The sample of individuals who had BASC-3s included 227 individuals; however, after excluding individuals based on whether they had valid BASC and MMPI profiles, the sample was reduced to 175 valid resident profiles. Residents with valid BASC profiles ages ranged from 13 to 18 years of age ($M = 15.29$, $SD = 1.18$) and were

primarily male ($n = 116$) and White ($n = 140$, 80%). The valid MINI-KID sample included 333 residential profiles with the same demographics discussed in the above paragraph. However, samples of those who have specific modules varied based on the needs of the resident. As a result, module samples varied from 104 residents (MINI-KID Specific Phobia Module) to 122 (MINI-KID Depression Module).

Measures

Minnesota Multiphasic Personality Inventory-Adolescent-Restructured Form (MMPI-A-RF; Archer et al., 2016). The MMPI-A-RF is the most recent version of the MMPI designed for use with adolescents. Consisting of 241 true/false items, the MMPI-A-RF is a broad band instrument intended to assess the psychological functioning of adolescents ranging from 14 to 18 years old. Archer (2016) indicated the instrument can be used with youth as young as 12-years old, so long as they have the reading and comprehension skills necessary to complete the instrument.

The focus of the current study is the MMPI-A-RF Restructured Clinical scales, which are the core of the measure because they encapsulate nine key psychological constructs. Scale descriptions are provided in Table 1, and reliability and validity information has been discussed briefly above. However, more detailed scale descriptions are provided below.

Demoralization (RCd) scale. The RCd scale consists of 18-items and measures “unhappiness, poor morale” and life dissatisfaction (Archer, 2016). Adolescents who score high on RCd may experience a number of emotionally distressing factors such as sadness, depressed mood, low self-esteem, inattention, low energy, fatigue, and possibly suicidal ideation. High scores are also supposed reflect helplessness and hopelessness. Low scores, on the other hand, indicate a relatively high degree of life satisfaction and morale. Research conducted by Archer et

al. (2016) indicated that scores on RCd demonstrated adequate test-retest and internal consistency reliability in the MMPI-A-RF normative sample ($r = .82$; $\alpha = .80$ to $.83$) and inpatient treatment samples ($\alpha = .87$ to $.88$). Internal consistency reliability in the current study's sample of residents with valid MMPI-A-RF profiles was $\alpha = .91$. Archer et al. (2016) also found that RCd was significantly correlated to aspects of emotional distress but not markers of externalization, consistent with the construct of demoralization the scale is supposed to be measuring.

Somatic Complaints (RC1) scale. The 23-item RC1 scale measures a wide range of somatic complaints, such as head pains and gastrointestinal issues. Archer (2016) reported moderate elevations on RC1 reflect the report of physical health problems, and that higher scores may be linked to complaints that are the result of psychological preoccupation. Adolescents that score high on RC1 may report frequent somatic complaints related to head pains, gastrointestinal difficulties, problems concentrating, and/or low energy and fatigue. Archer et al. (2016) found that RC1 has strong positive relations with other measures of somatic complaints and, to a lesser degree, with feelings of anxiety and depression. Scores on RC1 demonstrated adequate test-retest and internal consistency reliability in the MMPI-A-RF normative sample ($r = .78$; $\alpha = .77$ to $.78$) and inpatient treatment samples ($\alpha = .83$ to $.84$) (Archer et al., 2016). Internal consistency reliability in the current sample of residents with valid MMPI-A-RF profiles was $\alpha = .76$. Archer et al. (2016) did not find strong relations with measures of externalization or thought dysfunction, supporting its scores' discriminant validity.

Low Positive Emotions (RC2) scale. RC2, a 23-item scale measuring a lack of positive emotional experiences, is similar to RCd in that elevated scores are related to feelings of hopelessness, sadness, depression, anxiety, and low self-esteem (Archer, 2016; Archer et al.,

2016). However, high scores are also related to feelings of inadequacy and inefficiency, as well as anhedonia and psychomotor retardation (Archer, 2016). Adolescents who score high on RC2 are theorized to be more likely to be socially withdrawn, introverted, self-degrading, difficult to motivate, and self-punishing (Archer, 2016). These associations coincide with what RC2 is supposed to measure, and there is research to support its discriminant validity in that Archer et al. (2016) found that RC2 had no meaningful associations with measures of externalization or thought dysfunction. Scores on RC2 demonstrated moderately adequate test-retest and internal consistency reliability in the MMPI-A-RF normative sample ($r = .65$; $\alpha = .53$ to $.60$) and inpatient treatment samples ($\alpha = .66$ to $.72$). A similar level of internal consistency was observed in the current study ($\alpha = .61$).

Cynicism (RC3) scale. RC3 consists of nine items and assesses negative and cynical views of interpersonal relationships. Elevated scores on RC3 have been related to cynical beliefs about others, distrustfulness, and engaging in rule-breaking behaviors (e.g., suspension, unlawful behavior) (Archer, 2016; Archer et al., 2016). It is theorized that adolescents who score high on RC3 may feel that the actions of others are not well-intentioned (Archer, 2016). Low scores, on the other hand, are theorized to reflect trust and positive views of others and their intentions (Archer, 2016). However, as previously mentioned, Archer et al. (2016) did not include direct measures of cynicism as external criteria, so there is a need for additional research to establish RC3 as a measure of cynicism (Graham et al., 2022). Scores on RC3 demonstrated adequate test-retest and internal consistency reliability in the MMPI-A-RF normative sample ($r = .60$; $\alpha = .77$ to $.80$) and inpatient treatment samples ($\alpha = .61$ to $.64$). The results of the current study provide further support that RC3 demonstrates adequate internal consistency reliability, as in the sample of residents with valid MMPI-A-RF profiles I observed $\alpha = .65$.

Antisocial Behavior (RC4) scale. Consisting of 20-items, RC4 assesses various antisocial and conduct disorder related behaviors. This scale was intended to measure disinhibited rule breaking and antisociality. Elevated scores on RC4 are related to being in trouble both at school and at home, hanging around socially undesirable peer groups, and a history of substance abuse (Archer, 2016; Archer et al., 2016). Low scores were connected with reduced risks for various acting-out behaviors (Archer, 2016). Adolescents with high scores on RC4 would be more likely to have a history of juvenile detention placements, criminal charges related to alcohol or drug use, running away from home, and school suspension (Archer, 2016). High scores on RC4 have also been linked to greater verbal and physical aggression towards others (Archer et al., 2016). There is also support for its discriminant validity in that Archer et al. (2016) found no consistent associations between RC4 and internalizing or thought dysfunction factors. Scores on RC4 demonstrated adequate test-retest and internal consistency reliability in the MMPI-A-RF normative sample ($r = .66$; $\alpha = .77$ to $.80$) and inpatient treatment samples ($\alpha = .83$ to $.85$). Internal consistency reliability in the current study's sample of residents with valid MMPI-A-RF profiles was $\alpha = .82$.

Ideas of Persecution (RC6) scale. RC6 includes nine items that measure persecutory beliefs that can rise to the level of paranoid delusions. Elevated scores on RC6 have been theorized to be related to feelings of mistrust, suspiciousness, and distrust as well as persecutory thinking, hallucinations, and psychotic symptoms (Archer, 2016). Archer et al. (2016) also suggest that youth experiencing ideas of persecution would have interpersonal difficulties with their peers. Archer et al. (2016) found that RC6 demonstrated small, yet consistent, associations with measures of emotional distress, anxiety, and problems with thinking. However, Archer et al. (2016) did not find strong evidence to support that RC6 measures persecutory ideation. Instead,

they demonstrated that adolescents that scored high on RC6 were more likely to engage in aggressive and oppositional behaviors. Graham et al. (2022) suggest that this may be due to the low bases rate of persecutory thinking and related difficulties in adolescents and because there was a lack of external criteria that adequately assesses persecutory ideation in the initial validation efforts. In comparison to other RC scales, scores on RC6 demonstrated slightly lower test-retest reliability but comparably adequate internal consistency reliability in the MMPI-A-RF normative sample ($r = .58$; $\alpha = .64$ to $.66$) and inpatient treatment samples ($\alpha = .71$ to $.73$) (Archer et al., 2016). The results of the current study provide further support that RC6 demonstrates adequate internal consistency reliability, as in the sample of residents with valid MMPI-A-RF profiles, $\alpha = .70$.

Dysfunctional Negative Emotions (RC7) scale. Consisting of 11-items, RC7 is a measure of negative emotional experiences related to “anxiety, irritability, impatience, apprehensiveness” and embarrassment (Archer, 2016). Adolescents with elevated scores on this scale may experience anxiety, many or specific fears, low-self-esteem, attentional and concentration difficulties, irritability, and impatience. High scores on RC7 are also likely to indicate anxiety-related difficulties, such as intrusive thoughts and nightmares (Graham et al., 2022). Elevated scores on this scale are also correlated with suicidal ideation and insecurity (Archer, 2016). Archer et al. (2016) found that RC7 has positive associations with a number of distress markers and depressed mood, supporting its convergent validity. Supporting RC7 scores’ discriminant validity, Archer et al. (2016) did not find meaningful associations with measures of externalizing and thought dysfunction. Scores on RC7 demonstrated adequate test-retest and internal consistency reliability in the MMPI-A-RF normative sample ($r = .74$; $\alpha = .63$ to $.66$) and

inpatient treatment samples ($\alpha = .63$ to $.78$) (Archer et al., 2016). Internal consistency reliability in the current study's sample of residents with valid MMPI-A-RF profiles was $\alpha = .80$.

Aberrant Experiences (RC8) scale. RC8 consists of eight items related to unusual thoughts and perceptions. Similar to RC6, RC8 is positively associated with unusual thinking and perceptions (Archer et al., 2016). However, RC8 has greater associations with auditory and visual hallucinations, as well as clinically significant psychoticism (Archer et al., 2016). Elevated scores on this scale have been related to difficulties with perceiving reality, disorientation, and day dreaming (Archer, 2016). Scores on this scale also demonstrated small relations with markers of social isolation, withdrawal, anxiousness, and depressed feelings (Archer et al., 2016). Scores on RC8 were not found to be associated with specific aspects of internalizing difficulties or externalizing behaviors, supporting their discriminant validity. In comparison to the other RC scales, scores on RC8 demonstrated slightly lower test-retest and internal consistency reliability in the MMPI-A-RF normative sample ($r = .56$; $\alpha = .55$ to $.59$) but demonstrated adequate internal consistency reliability in the inpatient treatment samples ($\alpha = .69$ to $.73$). Internal consistency reliability in the current study's sample of residents with valid MMPI-A-RF profiles was $\alpha = .66$.

Hypomanic Activation (RC9) scale. RC9 consists of eight items measuring hypomanic activity related to excitement and sensation seeking, high psychomotor energy, racing thoughts, and periods of sleeplessness. Elevated scores may indicate increased risk-taking and thrill-seeking as well as heightened level of activation, a history of conduct behaviors, and heightened aggressive behaviors. The correlates provided by Archer et al. (2016) showcased little evidence supporting the convergent validity of RC9 due to the low base rate of those symptoms in the validation sample (Graham et al., 2022). However, RC9 was found to be associated with markers

of aggressive behaviors in the validation sample, and in the validation sample boys receiving treatment in an inpatient care, higher energy levels were found (Archer et al., 2016). Thus, evidence in support of the convergent validity of RC9 scores is mixed. However, this may be due to the low base rate of hypomanic activation in youth. Thus, Graham et al. (2022) suggested that in populations without an elevated prevalence of hypomania, RC9 may be assessing impulsivity and excitement seeking, as well as a “disinhibited externalizing way of relating to the world” (p. 500). Scores on RC9 demonstrated adequate test-retest reliability whereas internal consistency reliability was demonstrated to be relatively lower in the MMPI-A-RF normative sample ($r = .71$; $\alpha = .45$ to $.52$) and inpatient treatment samples ($\alpha = .58$) in comparison to other RC scales. Similar internal consistency reliability was found for the current study in that RC9 demonstrated only moderately adequate internal consistency reliability in the sample of residents with valid MMPI-A-RF profiles ($\alpha = .57$).

Behavioral Assessment System for Children, 3rd Edition (BASC-3; Reynolds & Kamphaus, 2015). The BASC-3 is the most recent iteration of the BASC family of instruments. It is a multidimensional broad band family of instruments that use multiple methods to assess the psychological, emotional, cognitive, and school related difficulties experienced by children, adolescents, and young adults. There are five components of the BASC-3: the parent/caregiver rating scale (PRS), the teacher rating scale (TRS), a self-report scale of personality (SRP), a structured developmental history (SDH), and a student observation system (SOS). The current study will focus scales from SRP. For the SRP, test takers respond to over 100 items using a 4-point Likert-type scale (N = Never, S = Sometimes, O = Often, A = Almost Always). For descriptions of scales of the SRP, see Table 2.

There is some support for the reliability and validity of the SRP scales, as reported by Reynolds and Kamphaus (2015) in the BASC-3 manual. Specifically, for the SRP, test-retest reliability coefficients ranged from .72 (Test Anxiety) to .90 (Inattention/Hyperactivity), and combined α coefficients for adolescent boys and girls ranged from .77 (Test Anxiety and Locus of Control) to .96 (Internalizing Problems and Emotional Symptoms). With regards to validity, Reynolds and Kamphaus (2015) reported intercorrelation data for all of the scales of the SRP scales, and many of the scales showed strong correlations with conceptually relevant scales. The BASC-3 is also comparable to the BASC-2, and research regarding the validity of the BASC-2 has demonstrated moderate to large associations with other measures of problem behaviors in an adolescent community/outpatient sample (Mahan & Matson, 2011).

Mini International Neuropsychiatric interview for Children and Adolescents

Version 7.0 (MINI-KID; Sheehan, 2016). The MINI-KID is a structured diagnostic interview for DSM-5 psychiatric disorders in children and adolescents. It assesses for 30 of the most common clinically relevant disorders in pediatric mental health. There is a MINI-KID parent form and MINI-KID children and adolescent form; however, we will primarily use information gathered from the adolescent interview. It takes an average of 33 minutes to administer the MINI-KID, with questions being answered primarily using a “yes/no” format. Respondents are able to skip modules if they answer “no” to a screening question about a specific behavioral problem. The current study will use 16 of the modules that are relevant to the RC scales. For the list of the MINI-Kid modules that will be used, see Table 1.

There is some evidence to support the reliability and validity of the MINI-KID. Specifically, Duncan et al. (2018) found that test-retest reliability coefficients for the various modules ranged from .49 (Oppositional Defiant Disorder; ODD) to .79 (Alcohol

Abuse/Dependency and Anorexia/Bulimia). Duncan et al. (2018) also conducted confirmatory factor analyses, which provided evidence supporting the convergent and discriminant validity of the MINI-KID. Specifically, standardized factor loadings for the youth report ranged from .31 (Social Anxiety Disorder [SAD]; error variance = .90) to .93 (Generalized Anxiety Disorder [GAD]; error variance = .14).

Analyses

To investigate whether the MMPI-A-RF should include age stratified norms, the current study investigated whether age influences how the items comprise the construct assessed by the scale (i.e., measurement invariance) and how outcomes predicted by conceptually relevant RC scale scores (i.e., tests of moderation) are affected by age. Additional details regarding these analyses are below.

Measurement Invariance. In a sample of residents with valid MMPI-A-RF profiles ($n = 333$), I conducted confirmatory factor analyses (CFAs) for each of the RC scales without the addition of age to the model (i.e., a baseline model for each scale). This was done so that I could garner a better understanding of how the models were performing before considering age as a covariate. I then moved to examining measurement invariance. As previously mentioned, testing for measurement invariance would provide information regarding whether the same construct is being measured in the same way across age (Han et al., 2019). I could not use CFA alone, as procedures such as this require categorical grouping variables (Han et al., 2019) and my proposed moderator (i.e., age) were continuous. As such, it was best to use the Multiple Indicators Multiple Cause Model (MIMIC) to assess measurement invariance. This is because MIMIC allowed for continuous moderators in the assessment of measurement invariance (Woods et al., 2009).

Both the CFA and MIMIC model analyses were completed using M-Plus (Muthen & Muthen, 2017). In both cases, the Weighted Least Squares estimator was used to account for the measurement level of items (i.e., categorical item responses). I also used the same parameters for assessing whether the models were significant. Specifically, I used cut offs for global for normed fit indices as suggested by Sun (2005) (i.e., Comparative fit index [CFI] \geq .90, Tucker Lewis Index [TLI] \geq .90, Root mean square error of approximation [RMSEA] \leq .08, Standardized root mean square residual [SRMR], .08) and Yu (2002) (i.e., weighted root mean square residual [WRMR] close to 1). Local fit was assessed by whether factor loadings were statistically significant and greater than .30 (Nunnally, 1978). I also inspected the standardized parameter estimates to determine whether the demonstrated loadings were in the expected direction given the keying of the items.

Although there is not a general consensus for how to use MIMIC for measurement invariance testing, one approach was to conduct preliminary analyses to indicate which items are free of DIF. Specifically, each item was tested for DIF by constraining that item to equality in a model where all other items are presumed to be DIF-free (Woods et al., 2009). I compared a full model to the more constrained model created. Significant differences between the models identified using likelihood ratio (LR) difference tests for nested models indicated that an item had DIF (Woods et al., 2009). This process was repeated for every item contained in the model. Once this process was completed, a final MIMIC model was constructed for each scale and only items that showed significant DIF were regressed on our continuous variable (Woods et al., 2009). The final model provided group mean differences on the factor and DIF effects (Woods et al., 2009).

Predictive Validity. Before conducting the moderation analyses, I calculated initial baseline regression models for each of the outcome measures with the RC scales and age as predictors to better understand how the models fit without age as the moderator. To assess the potential differential validity of the MMPI-A-RF RC scales, I conducted a series of tests of moderation using regression. The predictor variable in each model were scores on an RC scale, while the outcome variable were scores on a conceptually relevant criterion measure. The outcomes examined in this study are outline in Table 1. In all models, the moderator variable was age. Essentially, the test of moderation examined whether the association between the MMPI-A-RF RC scale and the potential outcome measure changed as a function of age.

Initial regression models were calculated in SPSS (IBM Corp, 2021) using the Ordinary Least Squares (OLS) estimator because it provides a standard linear regression analysis for use with continuous variables. Overall, model fit was determined by significance of the coefficient of multiple correlation (i.e., R). Standardized regression coefficients (β) were used to evaluate the unique prediction offered by the RC scale scores and age. For moderation analyses, I used the Process Macro in SPSS (IBM Corp, 2021) to examine whether moderation occurred due to differences in the slope, intercept, or both. Variables included in the interaction were mean centered in all analyses. Overall model fit was determined by the significance of multiple R. Unstandardized regression coefficients for age, the RC scale, and the interaction between age and the RC scale (and their statistical significance) were evaluated to assess the unique predictions offered by the interaction between RC scale scores and age.

Results

Measurement Invariance

Baseline Models. Before investigating whether there was evidence of age-related item invariance, I examined the assumptions that needed to be met to conduct CFA and MIMIC analyses. I then calculated a CFA for each of the RC scales without adding age or investigating DIF to ascertain baseline model fit for each scale. Global model fit for each of these CFAs is presented in Table 3. Factor loadings for each of these models are presented in Appendix A.

Initial estimation of the models for RC1 and RC8 indicated the presence of a Heywood Case. For RC1, this involved Item 95 (Attacks of nausea), which had a very low base rate of endorsement in the keyed direction (26 True /333 False). This item's value was fixed to 1, and the model converged appropriately. For RC8, the Heywood Case involved Item 168 (See things others do not). This again appeared to be a problematic item due to a low base rate of endorsement in the keyed direction (29 True/333 False). Neither fixing the item's loading nor setting alternative start values resolved the problem. As such, the item was dropped from the model, and the revised model converged correctly.

As seen in Table 3, results of the CFAs indicated that RC3, RC8, and RC9 scales had good model fit, as the p -value associated with the models' chi-square estimates were greater than .05, RMSEA values were less than .05, CFI and TLI values were greater than or equal to .90, and WRMR values were close to 1. Although model chi-square values did not indicate good model fit for RCd, RC1, RC2, and RC7, other indices suggested good fit (RMSEA values less than .05, CFI and TLI values greater than or equal to .9, and WRMR close to 1). Results for RC6 indicated acceptable model fit as CFI and TLI were equal to .90, WRMR was 1.15, and the RMSEA was only slightly above .05. Results for RC4 indicated only acceptable model fit, potentially because of influence from items with low endorsement rates (e.g., Items 40 [Have not

gone to school when I should have; 26 True/333 False], 88 [Have been suspended; 36 True/333 False], and 160 [Have done bad things because friends; 66 True/333 False]).

As seen in Appendix A, when examining local fit, most items had statistically significant and meaningful ($\geq .3$; Nunally, 1978) loadings onto their respective factors. However, there were two items that did not significantly load onto to their scales. These included Item 78 (Hardly no heart pounding and short of breath; $p = .255$) on RC1 and Item 40 (Have not gone to school when should have; $p = .184$) on RC4. Additionally, there were three items that had significant, but low loadings on their respective factors. These included: 1) Item 163 (Enjoy play and recreation) on RC1, 2) Item 84 (Make up mind with ease) on RC2, and 3) Item 182 (Never done anything dangerous for thrill) on RC9. Finally, there were four items that loaded in the incorrect direction on RC1. These included Items 176 (Seldom have dizzy spell), 189 (Feel tight band around head), 192 (Stomach trouble), and 207 (Never bother by pain over heart). Overall, these results suggest the RC scales were functioning as they were designed, having good to acceptable global and local model fit. The exception was RC1, which had problems with local fit, thus, I suggest results of these and other analyses involving RC1 be viewed cautiously.

DIF Tests. After establishing baseline models for each RC scale, I then calculated a MIMIC model for each RC scale. The latent variable representing each RC scale was regressed onto age, and then each specific item from that scale was regressed onto age. Results for these models are in Tables 4 – 12, which present standardized parameter estimates describing the association between the latent variable representing the scale and age, as well as the DIF tests for all a scales' items.

RCd. As seen in Table 4, there were significant relations between RCd³ and age. As parameter estimates are negative, it would suggest that as age increases, standing on RCd decrease. DIF test results indicated that two items from RCd showcased significant DIF in the positive direction (i.e., Items 8 [Hard to keep mind on task] and 190 [Happy most of the time]). This indicates that as age increases the likelihood of responding false to the item increases.⁴

RC1. As seen in Table 5, there were no significant relations between RC1 and age. DIF test results indicated that two items from RC1 showcased significant DIF in the positive direction (i.e., Items 59 [Bothered by an upset stomach] and 187[Have been paralyzed or unusual weakness]). This indicates that as age increases the likelihood of responding false to these items increases.

RC2. As seen in Table 6, results of the current analysis for RC2 did not provide consistent evidence of an association between standing on RC2 and age. However, Items 169 (Expect to succeed) and 237 (No close friends) showcased significant, negative DIF, indicating that as age increases the likelihood of responding true to these item increases.

RC3. As seen in Table 7, none of the investigated relations between standing on RC3 and age were significant. However, four items showcased significant DIF: Items 24 (Most are honest because of fear of getting caught), 31 (People blame others for their mistakes), 152 (Takes argument to convince others of the truth), and 196 (Anyone would tell a lie to keep out of trouble). Items 24 and 31 had negative coefficients, which would suggest that as age increases the likelihood of responding to the item as true increases. Items 152 and 196 had positive

³ For ease of reading, these results name the scale, but actually reference the latent variable representing the scale modeled in the MIMIC analyses.

⁴ All MMPI-A-RF items are keyed such that 1 = True and 2 = False.

coefficients, which would suggest that as age increases the likelihood of responding false to the items increases.

RC4. As seen in Table 8, there was no significant association between RC4 and age. However, four items did demonstrate significant DIF: Items 106 (Get into fights when drinking), 144 (Better if all laws were thrown away), 235 (Enjoy marijuana), and 238 (Never run away from home). For Items 106 and 144, the coefficients were positive, meaning that as age increases the likelihood of responding false to the items increases. For Items 235 and 238, the coefficients were negative, meaning that as age increases the likelihood of responding true to the items increases.

RC6. As seen in Table 6, there was no significant association between standing on RC6 and age. Additionally, there were no significant indicators of DIF at the item level.

RC7. See Table 10 for parameter estimates between RC7 and age as well as DIF for each of the RC7 items. The parameter estimates investigating relations between standing on RC7 and age were not significant. However, one item, Item 153 (Feel anxiety about something or someone), showed negative significant DIF, suggesting that as age increases the likelihood of responding false to this item increases.

RC8. As seen in Table 11, there were no significant associations between standing on RC8 and age. Additionally, no items showcased significant DIF.

RC9. As seen in Table 12, the association between standing on RC9 and age was not significant. Additionally, no items showcased significant DIF.

Tests of Moderation

Before I investigated any moderation, I first conducted initial regression analyses for each of the RC scales and their corresponding criterion measures. I did this to examine the

relation between the RC scales and the proposed criterion measures, as to show whether age greatly impacts the model when it is introduced as a moderator. After conducting the initial regression models, I then conducted tests of moderation with age as the moderator using the Process MACRO in SPSS. Results for the initial and moderation models are presented in Tables 13 – 21. The left most columns present the initial model results with both unstandardized and standardized coefficients. The right most columns contains the results for the tests of moderation, and present only unstandardized coefficients.

RCd. As seen in Table 13, in the initial regression analyses with only age and RCd as predictor variables all models tested accounted for a significant proportion of the variability in the tested outcome variables. In all these models, RCd scores were significant predictors of the outcome measures. When moderation was examined, all tested models accounted for a significant portion of the variability in the outcomes. Consistent with the initial regression analyses, scores on RCd were significant predictors of the outcome measures, whereas age was not. Additionally, the interaction term did not significantly predict any of the outcome variables, suggesting that age did not significantly moderate any of the relations between scores on RCd and its corresponding outcome measures.

RC1. Moderation results for RC1 are presented in Table 14. In the initial regression analysis with age and RC1 as predictors of scores on BASC-3 Somatic Complaints, the overall model was statistically significant. Scores on RC1 were a significant predictor of scores on BASC-3 Somatic Complaints. However, when testing whether age moderates that association, the model was no longer statistically significant. Overall, this suggests that age did not significantly moderate the relation between scores on RC1 and scores on Somatic Complaints.

RC2. Moderation results for RC2 are presented in Table 15. In the initial regression analyses with age and RC2 as predictors of scores on BASC-3 Depression and past and current depression symptom counts from the MINI-KID, the overall models were statistically significant. Scores on RC2 were a significant predictor of scores on BASC-3 Depression and past and current depression symptom counts from the MINI-KID. When testing whether age moderates these associations, all the models were statistically significant. However, while RC2 predicted the outcome measures, age and the interaction term did not significantly predict scores on BASC-3 Depression or the past and current depression symptom counts from the MINI-KID. Overall, this suggests that age did not significantly moderate the relations between scores on RC2 and the depression-related outcome measures.

RC3. Moderation results for RC3 are presented in Table 16. In the initial regression analyses with age and RC3 as predictors of scores on BASC-3 Interpersonal Relations, BASC-3 Attitudes Towards Teachers, and BASC-3 Self-Reliance scales, the overall models were statistically significant for only BASC-3 Interpersonal Relations and BASC-3 Attitudes Towards Teachers. RC3 scores were significant predictors of BASC-3 Interpersonal Relations and BASC-3 Attitudes Towards Teachers. When testing whether age moderates these associations, the models for the BASC-3 Interpersonal Relations and Attitudes Towards Teachers were significant. Consistent with the initial regression analyses, scores on RC3 were significant predictors of the outcome measures, with the exception of BASC-3 Self-Reliance. However, neither age nor the interaction term were significant predictors of the outcome variables, suggesting that age did not significantly moderate any of the relations between scores on RC3 and its corresponding outcome measures.

RC4. Moderation results for RC4 are presented in Table 17. In the initial regression analyses with age and RC4 as predictors of the corresponding criterion measures, the overall models were statistically significant with the exception of the MINI-KID Conduct Disorder. Scores on RC4 were significant predictors of all of the other criterion measures. When testing whether age moderated the relations between RC4 and the corresponding criterion measures, all of the models were statistically significant. However, while RC4 scores were a predictor of the outcome measures, neither age nor the interaction term were, suggesting that age did not significantly moderate any of the relations between scores on RC4 and its corresponding outcome measures.

RC6. Moderation results for RC6 are presented in Table 18. In the initial regression analyses with age and RC6 as predictors of scores on BASC-3 Atypicality and BASC-3 Attitudes Towards Teachers, the overall model for BASC-3 Atypicality was statistically significant. Scores on RC6 were a significant predictor of scores on the BASC-3 Atypicality scale. When testing whether age moderates these associations, the models were statistically significant. However, while RC6 scores were a significant predictor of the outcome measures, age and the interaction terms did not significantly predict scores on BASC-3 Atypicality BASC-3 or BASC-3 Attitudes Towards Teachers. Overall, this suggests that age did not significantly moderate the relations between scores on RC6 and scores on BASC-3 Atypicality and Attitudes Towards Teachers.

RC7. Moderation results for RC7 are presented in Table 19. In the initial regression analyses with age and RC7 as predictors of scores on its corresponding outcome measures, the overall models were statistically significant. Scores on RC7 were a significant predictor of all the criterion measures. When testing whether age moderates these associations, the models were statistically significant, with the exception of the models for BASC-3 Anxiety and MINI-KID

Obsessive Compulsive Disorder. However, while scores on RC7 were a predictor of the outcome measures, age and the interaction terms did not significantly predict the criterion measures.

Overall, this suggests that age did not significantly moderate the relations between scores on RC7 and its corresponding criterion measures.

RC8. Moderation results for RC8 are presented in Table 20. In the initial regression analyses with age and RC8 as predictors of scores on its corresponding outcome measures, the overall models were statistically significant. Scores on RC8 were a significant predictor of all of the criterion measures. When testing whether age moderates these associations, all of the models were statistically significant, with the exception of the model for BASC-3 Atypicality. Scores on RC8 were significant predictors of the outcome measures. However, neither age nor the interaction term significantly predicted scores and symptom counts on the criterion measures. Overall, this suggests that age did not significantly moderate the relations between scores on RC8 and its corresponding criterion measures.

RC9. Moderation results for RC9 are presented in Table 21. In the initial regression analyses with age and RC9 as predictors of scores on its corresponding outcome measures, the overall models were statistically significant with the exception of the model predicting the count of current mania symptoms from the MINI-KID. Scores on RC9 were a significant predictor of all of the criterion measures in which the overall model was significant. When testing whether age moderates these associations, the models were statistically significant, with the exception of past mania symptom counts from the MINI-KID. However, while RC9 scores were a predictor of the outcome measure, neither age nor the interaction terms significantly predicted scores and symptom counts on the criterion measures. Overall, this suggests that age did not significantly moderate the relations between scores on RC9 and its corresponding criterion measures.

Discussion

The current study sought to investigate whether the MMPI-A-RF should use age-stratified norms. Specifically, I wanted to investigate whether there would be any item-or scale-level invariance or differential validity for RC scale scores as a result of the influence of age. Overall, the results of MI testing suggest that several items have DIF and standing on one scale, RCd, was significantly influenced by age. There was no evidence of moderation of the associations between RC scale scores and conceptually relevant outcomes by age in the differential predictive validity tests. These results suggest that there may be some items that function differently across ages, but these differences may not be enough to result in interpretive differences. The one exception may be RCd. These results may be a product of developmental differences; however, they may also reflect characteristics specific to the nature of the sample.

Measurement Invariance

The tests of measurement invariance suggested several items were functioning differently when the influence of age was considered. A review of these items' contents suggested they could be rationally grouped into several different domains. Specifically, Items 106 (Get into a fight when drinking), 152 (Takes argument to convince others of truth), 190 (Happy most of the time), and 238 (Never run away from home) appear to reflect aspects of youth's emotional functioning, while Items 8 (Hard to keep the mind on tasks), 24 (Most are honest because of fear), 144 (Better if all laws were thrown away), and 196 (Anyone would tell a lie to keep out of trouble) appear to reflect cognitive difficulties. Item 169 (Expect to succeed) appears to reflect age differences in self-esteem. Items 31 (People blame others for their mistakes), 153 (Feel anxiety about something or someone), and 237 (No close friends) may reflect social changes that can occur during adolescence. Lastly, Items 59 (Bothered by an upset stomach) and 187 (Have

been paralyzed or unusual weakness) may reflect developmentally appropriate somatic complaints that can occur during adolescence. The reasons age may have influenced each of these domains of items will be explored below.

Adolescence is a period characterized by changes in youth's emotional experiences. Specifically, adolescents move away from expressing more positive emotions, a trend exhibited in childhood, towards a heightened awareness and expression of negative emotions as they approach adulthood (Denham, 2019). I think the trend of expressing more negative emotions is reflected in the differential item functioning of Item 190 (Happy most of the time), as older adolescents were more likely to endorse the item as "False" compared to younger adolescents. Item 190 is a reverse keyed item scored on scales intended to assess a test taker's level of emotional distress (e.g., RCd). Thus, this differential functioning suggests this item may not be capturing younger adolescents' experiences with negative emotions as well as it does for older youth who have developed an awareness and ability to express their negative feelings.

Additionally, as discussed in the introduction, there are biological changes (i.e., pubertal maturation) that stimulate heightened emotional arousal during adolescence. The ability to down regulate these heightened emotions that comes with age is related to decreased amygdala activity and increased activity in lateral prefrontal regions (Martin & Ochsner, 2016). Items 106 (Get into a fight when drinking), 152 (Takes argument to convince others of truth), and 238 (Never run away from home), all of which demonstrated differential functioning across age, may reflect these developmentally appropriate differences in affective response. These items appear to represent possible decisions influenced by adolescents' heightened emotions. It may be that older adolescents were more likely to respond "False" to Items 106 and 152 because they have increased abilities for making decisions that are not as influenced by emotional arousal. For

example, they may have the ability to recognize that arguing may not always convince others, even if they feel strongly about what they are arguing about. A similar logic may apply to Item 238, as older adolescents were more likely to respond “True” to this item, reflecting an endorsement of not acting impulsively in the face of a strong emotion. Specifically, as older adolescents may possess a greater capacity for emotion regulation than younger adolescents, they may be able to reason through the costs and benefits of running away; whereas younger adolescents may be more emotionally reactive and, as a result, be more likely to run away in the face of an emotionally triggering event or situation. If replicated in future studies, the differential functioning of these items could suggest that younger adolescents are over-pathologized by the test’s scores (e.g., defiant), when, in reality, they are endorsing items in a manner consistent with their developmental stage.

The differential functioning observed for some of the other items in this study may reflect developmental differences in cognition. Cognitively, as adolescents develop they have increased top-down cognitive control, which allows them to exhibit a greater ability for focusing their attention to achieve a goal (Crone & Dahl, 2012). Greater attentional capacities may be the influential ability that led to the differential item functioning of Item 8 (Hard to keep mind on task), as older adolescents were more likely to endorse this item as “False” compared to younger adolescents. This endorsement difference may reflect actual differences in the youth’s attentional capacities across ages, as older adolescents should not have as much difficulty with attentional control or keeping their minds on tasks as younger adolescents do (Kadosh et al., 2014). Thus, this item could be over-pathologizing younger adolescents, who are responding to this item in a developmentally appropriate manner.

Another cognitive theme noticed was related to decision-making. Decision-making is influenced by several factors, including perspective-taking abilities. By age 14, many adolescents possess theory of mind (e.g., mentalizing); however, they make more errors than adults in social perspective taking (Dumontheil et al., 2010), suggesting that perspective taking does not fully mature until early adulthood. Development differences in the ability to engage in perspective taking may have influenced the differential functioning observed for Items 24 (Most are honest because of fear), 144 (Better if all laws were thrown away), and 196 (Anyone would tell a lie to keep out of trouble). For Item 24, older adolescents were more likely to endorse this item as “True” in comparison to younger adolescents, suggesting that they understand that many people would be honest due to fears of consequences. This is in juxtaposition to Item 196, in which older adolescents were more likely to endorse the item as “False”, suggesting that they do not believe that overall people would lie to keep out of trouble. These items could be related to perspective taking in the sense that younger adolescents will probably agree with the idea that others will lie for their own gain, as that is what they would do. However, the way in which older and younger adolescents responded to these items does not make for easy interpretation. Theoretically, if older adolescents believed that people would be honest because of fear of getting caught, they should have also believed that anyone would tell a lie to keep out of trouble. However, this relation was not seen in my results. For this reason, it is not clear whether these items over- or under-pathologize the experiences of younger and older adolescents.

For Item 144, my results suggested older adolescents were more likely to endorse this item as “False” in comparison to younger adolescents. This suggests older adolescents are less likely to believe that all laws should be discarded than younger adolescents, perhaps reflecting that older adolescents tend to be better at perspective-taking and can reason why laws are

important for the greater good. Thus, endorsements on Item 144 in the scored direction could be capturing developmentally appropriate differences in executive functioning, such as lower perspective-taking capabilities in younger adolescents. This would mean this item is potentially over-pathologizing younger adolescents when compared to a normative sample consisting of both younger and older adolescents, when in reality they are endorsing this item in a developmentally expected manner.

Although older adolescents engage in more adult-like decision-making than younger adolescents, they are still prone to increased engagement in risk-taking behaviors, as seen by the fact that risk taking in contexts with ambiguous consequences peaks during adolescence (Romer & Reyna, 2017). Specifically, there is some research that would suggest that engagement in risk taking behaviors increases throughout adolescence, with the highest reports of alcohol use, cigarette smoking, marijuana use, and delinquency occurring in the 12th grade (Hooshmand et al., 2012). This is reflected in the differential item functioning observed in this study for Item 235 (Enjoy marijuana), as older adolescents were more likely to endorse this item as “True” when compared to younger adolescents. This would mean this item is potentially over-pathologizing older adolescents when compared to a normative sample consisting of both younger and older adolescents, when in reality they are endorsing this item in a developmentally appropriate manner. In this case, problematic marijuana use would be best identified by comparing older adolescents to similarly aged adolescents.

There are also differences in self-esteem that occur in adolescence that may have influenced the differential functioning observed for some of the items in this study. Generally speaking, self-esteem gradually increases from adolescence into adulthood (Bleidorn et al., 2016). This increase in self-esteem can be seen partially in the differential item functioning of

Item 169 (Expect to succeed), in which older adolescents were more likely to endorse “True” in comparison to younger adolescents. This is relevant to self-esteem, as older adolescents are demonstrating that they have greater confidence in their ability to succeed. As a result, this item could be over-pathologizing the experiences of younger adolescents, who are responding in a manner fitting of their developmental standing.

Socially, as adolescents get older, their peer friendships begin to stabilize and they become more discerning in distinguishing between “friends” and “acquaintances” (Poulin & Chan, 2010). In other words, they have stricter criteria for defining someone as “friend,” resulting in smaller social groups. This idea may be reflected in the differential item functioning observed for Item 237 (No close friends), which older adolescents were more likely to endorse “True” when compared to younger adolescents. This is relevant to friendship acquisition because older adolescents may be more likely to lose older friends and gain fewer new friends, which could substantially reduce the number of close friendships that they have, whereas younger adolescents are more likely to claim having a large number of friends. Thus, endorsement of this item could be making older adolescents appear more socially withdrawn when compared to a normative sample consisting of both younger and older adolescents.

There was a theme of developmentally appropriate differential endorsement of some items reflecting somatic complaints. Specifically, I observed differential item functioning for Items 59 (Bothered by an upset stomach) and 187 (Have been paralyzed or unusual weakness), where older adolescents were more likely to endorse “False” in comparison to younger adolescents. This would suggest that older adolescents do not endorse upset stomachs and feelings of weakness as often as younger adolescents, which aligns with research suggesting younger children tend to report stomach aches and dizziness more often than older children

(Campo & Fritch, 1994). Thus, differential endorsement of these items could be capturing common developmental concerns for adolescents as a whole rather than somatic complaints that suggest an underlying pathological issue.

There were two items for which a developmental explanation was untenable. Specifically, Items 31 (People blame others for their mistakes) and 153 (Feel anxiety about something or someone) were more likely to be endorsed by older adolescents as “True”, potentially suggesting that they have generally more mistrustful views of others and their intentions. This is inconsistent with developmental perspectives on trust, which conceptualize this as a developmental task for younger children (Erikson, 1963). Instead, this trend of older adolescents endorsing more mistrustful views could be the result of the sample itself. Youth in residential care tend to experience disproportionately high rates of trauma exposure, including physical, emotional, and sexual abuse and neglect (Boel-Studt, 2017). Trauma exposure, especially from individuals close to the adolescent, is associated with lower levels of relational and general trust (Gobin & Freyd, 2014). Thus, it may be that older adolescents at the residential facility have had more opportunities for trauma exposure that would impact their abilities to trust others, which impacted their responses to Items 31 and 153.

Although I have posited these two items were influenced by the nature of the sample, it would be important to consider this for all the DIF results. The current study used adolescent reports from a residential facility for children and adolescents experiencing a number of emotional and behavioral difficulties. Adolescents from residential facilities such as the one sampled in the current study are at a higher risk of having been exposed to various traumatic experiences (e.g., abuse and neglect) and having tumultuous family relationships, as has been exemplified by the works of Connor and colleagues (2004) and Harr and colleagues (2013).

These experiences with trauma and negative family relationships could have a negative impact on these adolescents' socioemotional and psychological development, as pervasive and persistent child maltreatment has been linked to a number of mental health, interpersonal, behavioral, and physical difficulties (Negri et al., 2019). In addition, adolescents in a residential facility are not permitted the same freedoms as adolescents in other environments resulting in possible developmental delays (Julian & McCall, 2011). Specifically, adolescents in a residential facility may not be able to interact with their environment in a manner that would foster the development of specific tasks associated with their developmental period (e.g., the task of identity development). For these reasons, the results of the current study may not be representative of the adolescent population, but the results could be reflecting factors unique to the residential treatment sample used. For example, older adolescents were more likely to endorse true to Item 237 (No close friends) than younger adolescents. However, this finding may not be due to development differences between younger and older adolescents, but instead, this finding could be the result of older adolescents in a residential facility experiencing a lack of social support resulting from possible maltreatment. With that said, to investigate whether the results of the current study are the result of developmental differences between younger and older adolescents or a byproduct of the sample itself, future research should seek to replicate the current study in a non-residential sample of adolescents with various lived experiences.

There are other limitations for the DIF analyses that should be considered when interpreting the results. Namely, given that I conducted numerous DIF analyses for several scales, it was likely that I would find DIF for at least one item. This is because conducting multiple tests increases the possibility of finding significance (i.e., Type I error), even if there is none (Shaffer, 1995), leading to an increase probability of incorrectly identifying DIF where it is

not present. Secondly, with MIMIC being a relatively newer measurement invariance technique, there is no particular standard for determining an effect size for DIF that would be considered clinically, and not just statistically, meaningful. Wood et al. (2009) used Benjamini-Hochberg's (1995) false discovery rate technique; however, Woods and colleagues (2009) reported that this method over corrected for multiple analyses. Thus, although I demonstrated several statistical differences in endorsement across age, the practical effect of the DIF is unclear. It maybe that they were not practically meaningful, as the presence of DIF for many of these items did not result in statistically significant associations between the latent variable representing scale scores and age, which suggests items with differential functioning may not be influencing overall scale functioning.

As I cannot definitively conclude the practical effects of these DIF items, if these results replicate in a different sample, it is imperative that future research take additional steps toward understanding whether the results reflect differences between age and item functioning or an empirical finding that does not reflect the true relation between age and RC scale item functioning. One way to address this, could be through conducting an analysis in which age is transformed from a continuous variable to a categorical variable (e.g., younger adolescents and older adolescents). Researchers would then investigate whether younger or older adolescents who endorse a particular DIF item as "True" or "False" will have different scores on a specific criterion scales in comparison to each other. For example, my results indicated that older adolescents were more likely to respond "True" to Item 235 (Enjoy Marijuana). Researchers would categorize younger and older adolescents, administer the MMPI-A-RF to get their responses to the item, and then the participants would be given a measure of risk-taking. If my DIF results are accurate, older adolescents who respond "True" to Item 235 should have higher

scores on risk-taking in comparison to younger adolescents who responded “False” to Item 106 (Get into fights when drinking). Ultimately, these results would provide further support for the idea that Item 235 over- pathologizes the experiences of older adolescents, who are responding in a manner which could be expected of them based on their developmental stage. Similar tests should be conducted for the other items demonstrating DIF, as well.

One exception to age not influencing the overall scale was RCd. The latent variable representing RCd demonstrated consistent, significant, negative associations with age in the DIF analyses, suggesting that as age increases standing on RCd tends to decrease. However, while RCd was found to have two items with DIF, Items 8 (Hard to keep mind on task) and 190 (Happy most of the time), these items did not impact the relation between one’s standing on RCd and age. This was demonstrated when conducted post-hoc analyses in which Items 8 and 190 were removed (individually and jointly) from the estimated MIMIC models. This allowed me to determine whether these items impacted the relation between one’s standing on RCd and age. Results suggested there was still a significant negative relation between one’s standing on RCd and age when these items were not included in the equations. This suggests alternative explanations for this association are needed.

One potential explanation for the association between standing on the latent variable representing RCd and age is consistent with the developmental literature, in that, scores on RCd capture emotional distress. Although adolescence is characterized by increased emotional arousal compared to adulthood, it is most prominent in the early stages of adolescence and tends to decrease as adolescents age (Steinberg, 2005). Thus, these results would suggest that RCd is insensitive to developmentally appropriate arousal in younger adolescents, as it characterizes their endorsements of emotional distress as pathology, rather than a developmentally expected

expression of emotional arousal. To investigate whether the trends continue with regards to the influence of age on RCd, it would be important to conduct a study that compares adolescents age, their endorsements on RCd, and their endorsements on a corresponding criterion scale such as emotional arousal or emotional responsivity. Researchers would break age down into categories (e.g., younger and older adolescents), administer the MMPI-A-RF to participants, and administer a corresponding criterion measure (e.g., a measure of emotional arousal). If the results of my analysis are correct, younger adolescents who have high endorsements on RCd should have higher scores on a measure of emotional arousal in comparison to older adolescents who have lower endorsements on RCd.

Another explanation that should be considered, however, is that this association may be a product of the residential nature of the sample. As previously mentioned, there are risk factors associated with youth in a residential sample and there are factors unique to living in a residential facility that can have negative impacts on the developmental trajectory of adolescent residents. For these reasons, the results regarding RCd and age could be the result of factors specific to the sample instead of development. While it was found that younger adolescents had a higher standing on the latent trait representing RCd than older adolescents, this may not have been the result of younger adolescents having higher emotional arousal than older adolescents. Instead, this trend could be the result of age-based differences in reasons for referral to the facility. Specifically, older adolescents may be referred to the facility for more externalizing difficulties related to law breaking whereas younger adolescents may be referred to the facility for experiencing intense emotional difficulties. This could explain why younger adolescents were found to have a higher standing on the latent trait representing RCd, a measure of emotional distress, in comparison to older adolescents. I did not have sufficient data to explore this

possibility in this study. As such, to determine if this is a sample specific finding, future research should seek to replicate the current study in a non-residential facility that ensure generalizability.

Another noteworthy caution for the DIF tests is related to Item 168 (Seeing things that others could not), which is scored on RC8. In initial analyses, Item 168 was a Heywood case, in which the standardized estimate was greater than 1. To address this issue, I attempted to set the value to 1 to allow the model to converge. However, the model still did not converge appropriately. As a result, Item 168 was removed from the model, and all the subsequent MI and DIF tests were calculated without this item. This problem may have been a result of Item 168 having a low base rate of endorsement, but, regardless, RC8 analyses do not represent the functioning of the scale as it is standardly scored. As such, future research should be conducted in a sample with a higher base rate of the type of difficulties this item captures, as to garner a holistic understanding of the item and RC8's functioning.

Although, in general, the scale models were functioning as intended by their developers (Archer et al., 2016), some items showed low item loadings, non-significant loadings, or loadings in the wrong direction. Items 84 (Make up mind with ease) scored on RC2, 33 (Never in trouble with the law) scored on RC4, and 182 (Never done anything dangerous for thrill) scored on RC9 showcased low item loadings of less than .30, which would suggest that the items themselves are not adequately contributing to their respective models (Nunally, 1978). Items 78 (Feels useless) scored on RC1 and 33 (Never in trouble with the law) scored on RC4 had nonsignificant *p*-values, suggesting that the items were adding error to their factor. Lastly, the parameters for Items 176 (Seldom have dizzy spells), 189 (Fell tight band around head), 192 (Stomach trouble), and 207 (Never bother by pain over heart) of RC1 were not in the keyed direction.

The previously mentioned item function issues of the current study can primarily be explained by difficulties within the sample. Data for this study were obtained from a residential treatment facility, specialized in working with children and adolescents involved with the juvenile justice system. This means there was a high base rate of some externalizing (e.g., Conduct Disorder) and internalizing phenomena (e.g., PTSD), but not others (e.g., Psychosis). Because of the inconsistent base rates for these items, I could not garner accurate parameter estimates regarding item functioning for many of the scales. Thus, future research should replicate this study in a larger more diverse sample in which there are greater base rates of a variety of presenting problems. This would provide greater power for the tests of moderation (discussed later), as well as give us more confidence in the MI results for scales in which base rate may have been an issue in this study (e.g., RC1).

As discussed in the previous paragraph, many of the problematic items in the MI analyses were scored on RC1. From inspecting the items that were not performing as would have been expected, I observed that Items 176 (Seldom have dizzy spells; 63 True/ 270 False) and 207 (Never bother by pain over heart; 52 True/ 281 False) showed a marked low overall endorsement of these items in the reverse scored direction. As well, Item 192 (Stomach trouble; 252 True/ 81 false) indicated a high overall endorsement of this difficulty in this sample Item 78 (Feels useless; 149 True/ 184 False) showed the opposite trend in which endorsements were variable. Additionally, all of these items' factor loadings were not in the adequate direction. As there was not a specific trend linking them all to explain these factor loadings, it may be the result of a problem with the scale itself. Although certainly we would want to conduct additional research on this issue, it may be that RC1 is an especially problematic scale to use with adolescents. RC1 was adapted from the adult version, the MMPI-2-RF. However, a developmental perspective was

not taken in the development of the scale. Instead, items from the MMPI-2-RF were maintained and included on the RC1 scale of the MMPI-A-RF if factor analyses suggested that the item significantly added onto the factor or model (Archer et al., 2016). As a result, the language that is used for some of the items may not be appropriate for an adolescent sample, making it likely that the scale is not capturing the way in which adolescents describe their somatic difficulties.

Therefore, we cannot be certain that the way in which adolescents are responding to these items is in coordination with how we believe the items should be interpreted.

Differential Predictive Validity

Although initial regression models suggested RC scale scores were significant predictors of most of the selected criterion measures, results of the moderation analyses suggested age did not moderate these relations. Once the interactions of age and RC scores were entered into the models, RC scale scores, age, and the interactions were not statistically significant predictors in any of the tested models. Thus, the results of the current analyses provide no evidence to support the idea that there is a need for age-based interpretations for the RC scales.

This relates to the DIF results that suggested that there may not be a need for age-based interpretations although there may be minor differences in item functioning. There are several issues that may explain these results. One, with the exception of RC3 and RC4 that had four items with DIF, many of the RC scales (e.g., RC7) only had one or two items that indicated DIF. As a result, there were not enough items indicating DIF to influence scores on the respective RC scale. However, RCd was an exception, as RCd only had two items with DIF and was consistently influenced by age. Two, for scales with more than two items with DIF, item loadings could have cancelled each other out. RC4 is an example of this, in that, Items 106 (Get into fights when drinking) and 144 (Better if all laws were thrown away) have positive parameter

estimates, and Items 235 (Enjoy marijuana) and 238 (Never run away from home) have negative parameter estimates.

Ultimately, just because the DIF items do not seem to be influencing RC scale scores, this does not mean it would not be important to investigate whether these same DIF items influence SP scale scores. While many of the items that showed DIF did not greatly impact the relations between scores on the RC scales and age, these items could still be impacting overall conclusions and possible scores on the SP scales. The RC scale items may be used in the SP scales, which can be shorter as they focus on specific problems. However, because they are shorter the presence of DIF items could have a greater impact on the scale score. For example, RC4 and the SP scale for Substance Abuse, a 4-item scale, share Item 235 (Enjoy marijuana), which results indicate has DIF. Because the substance abuse scale is so short, having an item with DIF, could greatly impact the relation between scores on its scale and age. As a result, it would be important to investigate the relations between specific DIF item endorsements, criterion measures reflecting their latent construct, and age to investigate whether age will impact scores on that specific criterion measure.

One of the primary limitations of the differential validity tests was that criteria selection was limited by the archival nature of the data used in this study. For example, the data lacked narrow criteria reflecting specific manifestations of psychotic phenomena that were appropriate for RC6 and RC8. I used the BASC-3 Atypicality scale, a broad scale reflecting odd behavior commonly associated with psychosis, as a criterion for both RC6 and RC8. However, these scales measure different phenomena, and thus, predictive validity differences may have been masked by the non-specific nature of the criterion measure. To address this concern, I conducted a post hoc factor analysis using items from the BASC-3 Atypicality scale to determine if more

specific facets of psychotic experience could be recovered from the scale's items. Results suggested three major factors arise: compulsive thinking, persecutory thinking, and hallucinations. After conducting this factor analysis, I conducted initial regression analyses with age and RC6 as the predictor variable for persecutory thinking and RC8 as a predictor of hallucinations. I then conducted tests of moderation for both of the RC scales. Overall, results provided further evidence that age is not a significant influence on scores on RC6 or RC8, as tests of moderation yielded non-significant results (unstandardized coefficient = .17 [RC6] and 2.41[RC8], $p = .85$ [RC6] and $.19$ [RC8], for RC6 and RC8 analyses, respectively). Nonetheless, this was to be expected as I used a short scale (i.e., BASC-3 Atypicality scale) with a low base rate and made it shorter by splitting it into subscales. Given this was an ad hoc solution to parsing variability in the Atypicality scale, it is important to remember that these are unvalidated outcome measures and to view these results cautiously. As such, future studies should continue examining the potential for differential validity using better criterion measures for these scales, likely in a sample where the base rates of these types of difficulties are higher, as well.

Post hoc solutions to the criteria problem as described above were not available for other scales, such as RC1 and RC3. RC1 is a measure of somatic complaints, and while I was able to use the BASC-3 Somatic Complaints scale, there was no interview data applicable for use. As a result, I could not use a measure of a different modality to circumvent the fact that similar scale modalities are likely to be related (Tehseen et al., 2017). Additionally, these data did not include a direct measure of cynicism, which would have been the best outcome measure for RC3 scores. Previous studies have found it difficult to narrow down RC3 scores' associations with other criterion measures (e.g., Sharf & Rogers, 2022); however, small associations have been made with scores on RC3 and rule breaking behaviors, mistrust, and self-interest (Archer, 2016,

Graham et al., 2022). For this reason, I used scales assessing the adolescent's relationships with others as RC3 criterion. However, Cynicism, as measured by the MMPI-A-RF, is supposed to measure adolescents who maintain a negative and cynical view of interpersonal relationships. As such, future research should replicate this study with criterion measures that best represent and coordinates with all the RC scales.

It would also be important to consider that the moderation analyses could have been underpowered to detect small effects, given that tests of moderation conducted within the regression context require very large sample sizes due to the introduction of a highly-correlated interaction term in the model (Shieh, 2009). Essentially, the distributional properties of the predictor and the moderator necessitates a large sample (Shieh, 2009). With that said, the current study did not have a large enough sample to adequately address the distributional nature of a moderation analysis. After my exclusionary process, the moderation sample with valid responses to the BASC-3 was reduced to a sample of 175, and the moderation sample for the MINI-KID modules vastly varied, with some analyses having 104 individuals while others had 122 individuals. Considering this limitation, the current study may not have been adequately powered to find significant interactions between the RC scales and age with relation to their corresponding outcome measures. As a result, having underpowered analyses makes it difficult to confidently conclude that the results of the current study provide strong evidence that age does not moderate the relation between RC scores and scores and symptom counts of the criterion measures.

Lastly, although I included both a self- and interview-based assessment of these youth's difficulties, an additional limitation of the current study is that both of these measures are based on the adolescent's report. Research has established that adult caregivers and youth do not

provide redundant diagnostic information (Klein, 1991), but instead they each provide a unique perspective into the adolescent's psychological well-being and emotional functioning. This is relevant, as it is not just that they report higher levels of different symptoms, but parents and children disagree about the presence or absence of externalizing difficulties, with generally more agreement for internalizing difficulties (Stanger & Lewis, 1993). Children, in general, report more affective symptoms than parents, whereas parents report greater conduct problems (Klein, 1991). However, children report more internalizing difficulties. As the current study investigated scales of both internalizing and externalizing difficulties, it would have been important to garner a better understanding of from both parents and children to get the most accurate information. No one informant's ratings can be used as a gold standard by which to measure psychopathology in children

Conclusion

The current study sought to investigate whether the MMPI-A-RF should use age-stratified norms. After analyzing the results of the current study, there was no significant evidence to support the hypotheses that there would be some evidence of measurement invariance and differential invalidity. Although there was some different item functioning found in some of the RC scales, these DIFs did not impact the relation between scores on the other RC scales and age. The tests of moderation did not provide any evidence to support any score differences as a result of age. The exception to this would be RCd, where results suggests that scale scores may over-pathologize the higher levels of emotional arousal experienced younger adolescents compared to older adolescents. If this result is replicated, test users should likely consider adopting age stratified norms or age specific interpretation strategies for this scale. However, results of the current study should be viewed cautiously, as this is the first study of its kind for

the MMPI-A-RF. There were issues with sample size and low base rates of item endorsement for some of the RC scales. It will also be important to replicate this work in a non-residential adolescent sample that is larger and more diverse with greater variability. This would aid in better understanding whether age is a significant moderator of the relation between scores on the RC scales and corresponding criterion measures, as well as it would help determine if the trends found in the current study was the result of development or a byproduct of factors specific to the sample. Future research should also seek to further assess the relations found in the current study to solidify the associations between RC scores and adolescent development.

References

- American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct* (2002, amended effective June 1, 2010, and January 1, 2017).
<https://www.apa.org/ethics/code/>
- Archer, R. P. (1984). Use of the MMPI with adolescents: A review of salient issues. *Clinical Psychology Review*, 4(3), 241–251. [https://doi.org/10.1016/0272-7358\(84\)90002-3](https://doi.org/10.1016/0272-7358(84)90002-3)
- Archer, R.P. (2016). *Assessing Adolescent Psychopathology: MMPI-A / MMPI-A-RF, Fourth Edition* (4th ed.). Routledge. <https://doi.org/10.4324/9781315737010>
- Archer, R. P., Stolberg, A. L., Gordon, R. A., & Goldman, W. R. (1986). Parent and child MMPI responses: Characteristics among families with adolescents in inpatient and outpatient settings. *Journal of Abnormal Child Psychology*, 14(1), 181–190. <https://doi.org/10.1007/BF00917232>
- Archer, R. P., Handel, R. W., Ben-Porath, Y. S., & Tellegen, A. (2016). *Minnesota Multiphasic Personality Inventory-Adolescent-Restructured Form (MMPI-A-RF): Manual for Administration, Scoring, Interpretation, and Technical Manual*. Minneapolis: University of Minnesota Press.
- Archer, R. P., Maruish, M., Imhof, E. A., & Piotrowski, C. (1991). Psychological test usage with adolescent clients: 1990 survey findings. *Professional Psychology: Research and Practice*, 22(3), 247–252. <https://doi.org/10.1037/0735-7028.22.3.247>
- Archer, R. P., & Newsom, C. R. (2000). Psychological test usage with adolescent clients: Survey update. *Assessment*, 7(3), 227-235. <https://doi.org/10.1177/107319110000700303>
- Arain, M., Haque, M., Johal, L., Mathur, P., Nel, W., Rais, A., ... & Sharma, S. (2013). Maturation of the adolescent brain. *Neuropsychiatric disease and treatment*, 449-461.

- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Prentice Hall/Pearson Education.
- Backes, E. P., Bonnie, R. J., & National Academies of Sciences, Engineering, and Medicine. (2019). Adolescent Development. *The Promise of Adolescence: Realizing Opportunity for All Youth*. National Academies Press (US).
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57, 289–300.
- Ben-Porath, Y. S., & Tellegen, A. (2011). *The Minnesota Multiphasic Personality Inventory2-Restructured Form (MMPI-2-RF): Manual for administration, scoring, and interpretation*. University of Minnesota Press. Original work published in 2008.
- Benson, N., & Keith, T. Z. (2017). Wechsler Intelligence Scale for Children--Fifth Edition. *The Twentieth Mental Measurements Yearbook*. Berenbaum, S. A., Beltz, A. M., & Corley, R. (2015). The importance of puberty for adolescent development: conceptualization and measurement. *Advances in child development and behavior*, 48, 53-92.
- Bleidorn, W., Arslan, R. C., Denissen, J. J. A., Rentfrow, P. J., Gebauer, J. E., Potter, J., & Gosling, S. D. (2016). Age and gender differences in self-esteem—A cross-cultural window. *Journal of Personality and Social Psychology*, 111(3), 396–410.
doi:10.1037/pspp0000078
- Boel-Studt, S. M. (2017). Latent subtypes of youth in psychiatric residential care. *Children and Youth Services Review*, 77, 76-85.

- Butcher, J. N., & Pope, K. S. (1992). The research base, psychometric properties, and clinical uses of the MMPI-2 and MMPI-A. *Canadian Psychology / Psychologie canadienne*, 33(1), 61–78. <https://doi.org/10.1037/h0078693>
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A. M., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring*. Minneapolis: University of Minnesota Press.
- Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R. P., Tellegen, A., Ben-Porath, Y. S., & Kaemmer, B. (1992). *MMPI-A: Minnesota Multiphasic Personality Inventory—Adolescent: Manual for administration, scoring, and interpretation*. University of Minnesota Press
- Campo, J. V., & Fritsch, S. L. (1994). Somatization in Children and Adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry*, 33(9), 1223–1235. doi:10.1097/00004583-199411000-00003
- Cashel, M. L. (2002). Child and adolescent psychological assessment: Current clinical practices and the impact of managed care. *Professional Psychology: Research and Practice*, 33(5), 446–453. <https://doi.org/10.1037/0735-7028.33.5.446>
- Committee on Psychological Testing, I. V. T., Populations, B. on the H. of S., & Medicine, I. of. (2015). Overview of Psychological Testing. In *Psychological Testing in the Service of Disability Determination*. National Academies Press (US). <https://www.ncbi.nlm.nih.gov/books/NBK305233/>
- Connor, D. F., Doerfler, L. A., Toscano, P. F., Volungis, A. M., & Steingard, R. J. (2004). Characteristics of children and adolescents admitted to a residential treatment center. *Journal of Child and Family Studies*, 13, 497-510.

- Christie, D., & Viner, R. (2005). Adolescent development. *BMJ*, *330*(7486), 301–304.
doi:10.1136/bmj.330.7486.301
- Crone, E. A., & Dahl, R. E. (2012). Understanding adolescence as a period of social–affective engagement and goal flexibility. *Nature Reviews Neuroscience*, *13*(9), 636–650. doi:10.1038/nrn3313
- Denham, S. A. (2019). Emotional Competence During Childhood and Adolescence. *Handbook of Emotional Development*, 493–541. doi:10.1007/978-3-030-17332-6_20
- Dumontheil, I., Küster, O., Apperly, I. A., & Blakemore, S.-J. (2010). Taking perspective into account in a communicative task. *Neuroimage*, *52*, 1574–1583.
doi: 10.1016/j.neuroimage.2010.05.056
- Erickson, E. (1963). *Childhood and society* (2nd ed.). New York, NY: Norton.
- Graham, J. R., Veltri, C. O. C., & Lee, T. T. C. (2022). *MMPI Instruments. Assessing Personality and Psychopathology* (6th ed.). Oxford University Press.
- Gobin, R. L., & Freyd, J. J. (2014). The impact of betrayal trauma on the tendency to trust. *Psychological Trauma: Theory, Research, Practice, and Policy*, *6*(5), 505.
- Han, K., Colarelli, S. M., & Weed, N. C. (2019). Methodological and statistical advances in the consideration of cultural diversity in assessment: A critical review of group classification and measurement invariance testing. *Psychological Assessment*, *31*(12), 1481–1496.
<https://doi.org/10.1037/pas0000731>
- Harkness, A. R., & McNulty, J. L. (1994). The Personality Psychopathology Five (PSY-5): Issues from the pages of a diagnostic manual instead of a dictionary. In S. Strack & M. Lorr (Eds.), *Differentiating normal and abnormal personality* (pp. 291–315). Springer Publishing Company.

- Harr, C. R., Horn-Johnson, T. C., Williams, N. J., Jones, M., & Riley, K. (2013). Personal trauma and risk behaviors among youth entering residential treatment. *Child and Adolescent Social Work Journal, 30*, 383-398.
- Hooshmand, S., Willoughby, T., & Good, M. (2011). Does the direction of effects in the association between depressive symptoms and health-risk behaviors differ by behavior? A longitudinal study across the high school years. *Journal of Adolescent Health, 50*(2), 140-147. <https://doi.org/10.1016/j.jadohealth.2011.05.016>
- IBM Corp. Released 2021. IBM SPSS Statistics for Windows, Version 28.0. Armonk, NY: IBM Corp
- Julian, M. M., & McCall, R. B. (2011). The development of children within alternative residential care environments. *International Journal of Child & Family Welfare, 14*(3-4), 119-147.
- Kadosh, K. C., Heathcote, L. C., & Lau, J. Y. (2014). Age-related changes in attentional control across adolescence: How does this impact emotion regulation capacities? *Frontiers in Psychology, 5*, 111.
- Klein, R. G. (1991). Parent-child agreement in clinical assessment of anxiety and other psychopathology: A review. *Journal of Anxiety Disorders, 5*(2), 187-198.
doi:10.1016/0887-6185(91)90028-r
- Klimstra, T., Kuppens, P., Luyckx, K., Branje, S., Hale, W., Oosterwegel, A., Koot, H., & Meeus, W. (2015). Daily dynamics of adolescent mood and identity. *Journal of Research on Adolescence, 26*(3), 459-473. doi:10.1111/jora.12205
- Klinge, V., Lachar, D., Grisell, J., & Berman, W. (1978). Effects of scoring norms on adolescent psychiatric drug users' and nonusers' MMPI profiles. *Adolescence, 13*(49), 1-11.

- Knoll, L. J., Magis-Weinberg, L., Speekenbrink, M., & Blakemore, S. J. (2015). Social Influence on Risk Perception During Adolescence. *Psychological Science*, 26(5), 583–592. doi:10.1177/0956797615569578
- Marks, P.A., Seeman, W., & Haller, D.L. (1974). *The actuarial use of the MMPI with adolescents and adults*. Baltimore: Williams & Wilkings.
- Martin, R. E., & Ochsner, K. N. (2016). The neuroscience of emotion regulation development: Implications for education. *Current opinion in behavioral sciences*, 10, 142-148.
- Morris, A.S., Silk, J.S., Steinberg, L., Myers, S.S., & Robinson, L.R. (2007). The role of the family context in the development of emotion regulation. *Social Development*, 16(2), 361–388. <https://doi.org/10.1111/j.1467-9507.2007.00389.x>
- Muthén, L.K. and Muthén, B.O. (2017). *Mplus User's Guide*. Eighth Edition. Los Angeles, CA: Muthén & Muthén
- Negriff, S., Gordis, E. B., Susman, E. J., Kim, K., Peckins, M. K., Schneiderman, J. U., & Mennen, F. E. (2019). The Young Adolescent Project: A longitudinal study of the effects of maltreatment on adolescent development. *Development and Psychopathology*, 1–20. doi:10.1017/s0954579419001391
- Peper, J. S., & Dahl, R. E. (2013). The teenage brain. *Current Directions in Psychological Science*, 22(2), 134–139. doi:10.1177/0963721412473755
- Poulin, F., & Chan, A. (2010). *Friendship stability and change in childhood and adolescence*. *Developmental Review*, 30(3), 257–272. doi:10.1016/j.dr.2009.01.001
- Reiter, E.O., & Root, A. W. (1975). Hormonal changes of adolescence. *Medical Clinics of North American*, 59(6), 1289 -1304. [https://doi.org/10.1016/S0025-7125\(16\)31930-7](https://doi.org/10.1016/S0025-7125(16)31930-7)

- Reynolds, C. R., & Kamphaus, R. W. (2015). *Behavior assessment system for children* (3rd ed.).
Bloomington: NCS Pearson, Inc. (BASC–3).
- Reynolds, C. R., & Suzuki, L. A. (2012). Bias in Psychological Assessment: An Empirical
Review and Recommendations. In I. Weiner (Ed.), *Handbook of Psychology, Second
Edition* (p. hop210004). John Wiley & Sons, Inc.
<https://doi.org/10.1002/9781118133880.hop210004h>
- Sanders, R. A. (2013). Adolescent psychosocial, social, and cognitive development. *Pediatrics in
Review, 34*(8), 354–359. doi:10.1542/pir.34-8-354
- Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and
implications. *Human Resource Management Review, 18*(4), 210–222.
<https://doi.org/10.1016/j.hrmr.2008.03.003>
- Sellbom, M. (2019). The MMPI-2-Restructured form (MMPI-2-RF): Assessment of personality
and psychopathology in the twenty-first century. *Annual Review of Clinical Psychology,*
15, 149–177. <https://doi.org/10.1146/annurev-clinpsy-050718-095701>
- Sharf, A. J., & Rogers, R. (2020). Validation of the MMPI-A-RF for youth with mental health
needs: A systematic examination of clinical correlates and construct validity. *Journal of
Psychopathology and Behavioral Assessment, 42*(3), 527-538.
<https://doi.org/10.1007/s10862-019-09754-x>
- Sheehan, D. V. (2016). *Mini International Neuropsychiatric Interview for Children and
Adolescents English version 7.0.2*. Jacksonville, FL: Medical Outcome Systems.
- Shieh, G. (2009). Detecting interaction effects in moderated multiple regression with continuous
variables power and sample size considerations. *Sage Journals, 12*(3), 510-528.
<https://doi.org/10.1177/1094428108320370>

- Smetana, J. G., Campione-Barr, N., & Metzger, A. (2006). Adolescent development in interpersonal and societal contexts. *Annual Review of Psychology*, *57*(1), 255–284. <https://doi.org/10.1146/annurev.psych.57.102904.190124>
- Stanger, C., & Lewis, M. (1993). Agreement among parents, teachers, and children on internalizing and externalizing behavior problems. *Journal of Clinical Child Psychology*, *22*(1), 107–116. doi:10.1207/s15374424jccp2201_11
- Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Developmental Review*, *28*(1), 78–106. doi:10.1016/j.dr.2007.08.002
- Steinberg, L. (2005). Cognitive and affective development in adolescence. *Trends in cognitive sciences*, *9*(2), 69-74. <https://doi.org/10.1016/j.tics.2004.12.005>
- Steinberg, L., & Morris, A. S. (2001). Adolescent development. *Journal of Cognitive Education and Psychology*, *2*(1), 55-87. <https://doi.org/10.1891/194589501787383444>
- Stokes, J. M., Pogge, D. L., & Archer, R. P. (2018). Comparisons between the Minnesota Multiphasic Personality Inventory-Adolescent-Restructured Form (MMPI-A RF) and MMPI-A in adolescent psychiatric inpatients. *Psychological Assessment*, *30*(3), 370–382. <https://doi.org/10.1037/pas0000488>
- Sun, J. (2005). Assessing goodness of fit in confirmatory factor analysis. *Measurement and Evaluation in Counseling and Development*, *37*(4), 240–256. doi:10.1080/07481756.2005.1190976
- Szczesniak, M., Colaco, M., & Rondon, G. (2012). Development of interpersonal trust among children and adolescents. *Polish Psychological Bulletin*, *43*(1), 50-58. <https://doi.org/10.2478/v10059-012-0006-5>
- Tehseen, S., Ramayah, T., & Sajilan, S. (2017). Testing and controlling for common method

- variance: A review of available methods. *Journal of Management Sciences*, 4(2), 146-175. DOI:10.20547/jms.2014.1704202
- Tellegen, A., & Ben-Porath, Y. S. (2011). *Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF): Technical manual*. University of Minnesota Press. Original work published in 2008.
- Tellegen, A., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., Graham, J. R., & Kaemmer, B. (2003). *MMPI-2 Restructured Clinical (RC) scales: Development, validation, and interpretation*. University of Minnesota Press.
- Wechsler, D. (2014). *WISC-V: Technical and Interpretive Manual*. Bloomington, MN: Pearson.
- Woods, C. M., Oltmanns, T. F., & Turkheimer, E. (2009). Illustration of MIMIC-model DIF testing with the schedule for nonadaptive and adaptive personality. *Journal of psychopathology and behavioral assessment*, 31, 320-330.
- Van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Editorial: Measurement invariance. *Frontiers in Psychology*, 6. <https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01064>
- Yu, C. (2002). Evaluating cutoff criteria of model fit indices for latent variable models with binary or continuous outcomes (UMI Number: 3066425) [Doctoral Thesis, University of California]. UMI Dissertation Publishing.

Table 1*RC Scales and Criterion Scales*

| RC Scales | | Description | Outcome Measures | |
|-----------|-----------------------|---|------------------|---|
| RCd | Demoralization | Measures dimensions of unhappiness, poor morale, and life dissatisfaction | BASC-3 | Depression Scale |
| | | | BASC-3 | Locus of Control Scale |
| | | | BASC-3 | Self-Esteem Scale |
| | | | BASC-3 | Sense of Inadequacy Scale |
| | | | MINI-KID | Major Depressive Episode |
| | | | MINI-KID | Suicidality & Suicide Behavior Disorder |
| RC 1 | Somatic Complaints | Assess for a wide range of somatic complaints | BASC-3 | Somatization Scale |
| RC 2 | Low Positive Emotions | Captures a lack of positive emotional experiences | BASC-3 | Depression Scale |
| | | | MINI-KID | Major Depressive Episode Disorder |
| RC 3 | Cynicism | Asses for negative views of interpersonal relationships | BASC-3 | Interpersonal Relations Scale |
| | | | BASC-3 | Attitudes to Teachers Scale |
| | | | BASC-3 | Self-Reliance Scale |
| RC 4 | Antisocial Behavior | Captures various antisocial and conduct disordered related behaviors | BASC-3 | Anger Control Scale |

| | | | | |
|------|---------------------------------|---|----------|---------------------------------|
| | | | BASC-3 | Alcohol Abuse Scale |
| | | | MINI-KID | Conduct Disorder |
| | | | MINI-KID | Oppositional Defiant |
| | | | MINI-KID | Alcohol Use Disorder |
| | | | MINI-KID | Other Substance Use Disorder |
| RC 6 | Ideas of Persecution | Measures persecutory beliefs to the point of paranoid delusions and beliefs | BASC-3 | Atypicality Scale |
| | | | BASC-3 | Attitudes Toward Teachers Scale |
| RC 7 | Dysfunctional Negative Emotions | Measures negative emotional experiences related to anxiety, irritability, impatience, apprehensiveness, and embarrassment | BASC-3 | Anxiety Scale |
| | | | MINI-KID | Generalized Anxiety Disorder |
| | | | MINI-KID | Panic Disorder |
| | | | MINI-KID | Agoraphobia |
| | | | MINI-KID | Separation Anxiety |
| | | | MINI-KID | Specific Phobia |
| | | | MINI-KID | Obsessive-Compulsive Disorder |
| | | | MINI-KID | Post-Traumatic Stress Disorder |
| RC 8 | Aberrant Experiences | Captures unusual thoughts and perceptions | BASC-3 | Atypicality Scale |

| | | | | |
|------|----------------------|---|----------|----------------------------------|
| | | | MINI Kid | Psychotic Symptoms and Disorders |
| RC 9 | Hypomanic Activation | Measures hypomanic activity related to excitement seeking, high energy, racing thoughts, and periods of sleeplessness | BASC-3 | Mania Scale |
| | | | BASC-3 | Sensation Seeking Scale |
| | | | BASC-3 | Hyperactivity Scale |
| | | | MINI-KID | Mania |

Note. BASC-3= Behavioral Assessment System for Children, Third Edition; MINI-KID= MINI International Neuropsychiatric Interview for Children and Adolescents, Version 7.02.

Table 2
BASC-3 Self-Report and Parent Report Scale Description

| Self-Report Scales | Description |
|-------------------------|--|
| Alcohol Abuse | The tendency to use alcohol to alleviate negative feelings and adverse outcomes as a result of alcohol use |
| Anxiety | Assesses the tendency to be nervous, fearful, or worried |
| Attention Problems | Measures the tendency to be easily distracted and unable to concentrate |
| Attitudes to Teachers | Feelings of resentment and dislike of teachers, beliefs that teachers are unfair, uncaring, or overly demanding. |
| Atypicality | Assess odd behavior commonly associated with psychosis. |
| Depression | Feelings of unhappiness, sadness, and dejection; a belief that nothing goes right |
| Hyperactivity | The tendency to be overly active, rush through work, and act without thinking |
| Interpersonal Relations | The perception of having good social relationships and friendships with peers. |
| Locus of Control | The belief that regards and punishments are controlled by external events or people |
| Self-Esteem | Feelings of self-esteem, self-respect, and self-acceptance |
| Self-Reliance Scale | Confidence in one's ability to solve problems; a belief in one's personal dependability and decisiveness |

| | |
|---------------------------|--|
| Sensation Seeking | The tendency to take risks and to seek excitement |
| Sense of Inadequacy Scale | Perceptions of being unsuccessful in school, unable to achieve one's goals, and general inadequate |
| Somatization | The tendency to be overly sensitive to, to experience, or to complain about relatively minor physical problems and discomforts |

Note. BASC-3= Behavioral Assessment System for Children, Third Edition.

Table 3*Goodness-of-Fit Indicators of Models for the Restructured Clinical Scales (n=333)*

| Model | χ^2 | p | df | CFI | TLI | RMSEA | RMSEA 90% CI | WRMR |
|-------|----------|------|------|-----|-----|-------|--------------|------|
| RCd | 191.93 | .001 | 135 | .99 | .99 | .04 | .02 -.05 | .86 |
| RC1 | 427.10 | .000 | 252 | .91 | .90 | .05 | .04-.05 | 1.14 |
| RC2 | 61.93 | .003 | 35 | .92 | .90 | .05 | .03 - .07 | .96 |
| RC3 | 36.43 | .106 | 27 | .97 | .96 | .03 | .00 - .06 | .79 |
| RC4 | 605.77 | .001 | 170 | .82 | .79 | .09 | .08 - .10 | 1.73 |
| RC6 | 77.87 | .001 | 27 | .92 | .90 | .08 | .06 - .10 | 1.15 |
| RC7 | 81.50 | .001 | 44 | .97 | .96 | .05 | .03 - .07 | .93 |
| RC8 | 20.18 | .125 | 14 | .98 | .95 | .04 | .00 - .07 | .74 |
| RC9 | 27.63 | .118 | 20 | .97 | .95 | .03 | .00 -.06 | .77 |

Note. df = degrees of freedom; CFI = comparative fit index, TLI =Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; RMSEA 90% CI = Root Mean Square Error of Approximation 90% Confidence Interval; WRMR = Weighted Root Mean Square Residual; p = probability; 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 4*Differential Item Functioning: RCd*

| Item | | RCd on Age | | | Item on Age | | |
|------|--|-------------|------------|-------------|-------------|------------|-------------|
| | | Estimate | S.E. | <i>p</i> | Estimate | S.E. | <i>p</i> |
| 8 | Hard to keep mind on task. | -.16 | .06 | .007 | .17 | .06 | .010 |
| 28 | Could not overcome difficulties. | -.14 | .06 | .013 | -.06 | .06 | .317 |
| 34 | Even when with people, feel lonely. | -.14 | .06 | .015 | -.07 | .06 | .222 |
| 60 | Feel helpless when making important decisions. | -.15 | .06 | .011 | .00 | .06 | .999 |
| 69 | Stand up for what is right. | -.15 | .06 | .011 | -.01 | .06 | .933 |
| 73 | Lacking self-confidence. | -.14 | .06 | .012 | -.03 | .06 | .658 |
| 74 | Don't care what happens. | -.15 | .06 | .010 | .03 | .07 | .643 |
| 79 | Feels useless. | -.14 | .06 | .011 | -.01 | .05 | .752 |
| 82 | Not happy with self. | -.15 | .06 | .010 | .02 | .06 | .757 |
| 102 | Difficulty starting things. | -.14 | .06 | .012 | -.04 | .06 | .492 |
| 128 | About to go to pieces. | -.14 | .06 | .012 | -.04 | .06 | .450 |
| 138 | Life is a strain. | -.15 | .06 | .008 | .08 | .06 | .138 |
| 162 | Not happy with life, it's too late. | -.15 | .06 | .009 | .07 | .06 | .311 |
| 170 | Something wrong with mind. | -.14 | .06 | .014 | -.08 | .07 | .248 |
| 183 | Wish to be happy. | -.14 | .06 | .012 | -.02 | .05 | .738 |
| 190 | Happy most of the time. * | -.14 | .06 | .016 | .12 | .06 | .045 |
| 230 | Everything is fine, I don't care. | -.15 | .06 | .010 | .03 | .06 | .623 |
| 234 | I am no good. | -.15 | .06 | .007 | .09 | .05 | .095 |

Note. Significant *p*-values are bolded. RCd = Demoralization; Estimate = Standardized Estimate; S.E. = Standard Error; *p* = probability. All items contribute to the total score. If they are endorsed true, with the exception of those marked with an *, which contribute to the total score when endorsed false.

Table 5*Differential Item Functioning: RC1*

| Item | | RC1 on Age | | | Item on Age | | |
|------|---|------------|------|----------|-------------|------------|-------------|
| | | Estimate | S.E. | <i>p</i> | Estimate | S.E. | <i>p</i> |
| 10 | No difficulties keeping balance. * | -.03 | .06 | .573 | -.06 | .07 | .398 |
| 20 | Lump in my throat. | -.03 | .06 | .678 | -.11 | .08 | .191 |
| 21 | Periods when couldn't do things because couldn't get going. | -.02 | .06 | .717 | -.10 | .06 | .108 |
| 22 | Head seems to hurt. | -.04 | .06 | .501 | .09 | .06 | .125 |
| 23 | Easily embarrassed. | -.03 | .06 | .687 | -.07 | 63.00 | .241 |
| 25 | Few or no pains. * | -.03 | .06 | .663 | .06 | .06 | .369 |
| 59 | Bothered by an upset stomach. | -.05 | .06 | .423 | .22 | .07 | .002 |
| 63 | Top of head feels tender. | -.03 | .06 | .666 | -.08 | .08 | .268 |
| 66 | Hardly feel pain the neck. * | -.03 | .06 | .623 | .02 | .07 | .816 |
| 78 | Hardly no heart pounding and short of breath. * | -.03 | .06 | .615 | .05 | .07 | .499 |
| 85 | No trouble with muscles twitching. * | -.03 | .06 | .644 | .05 | .07 | .469 |
| 95 | Attacks of nausea. | -.03 | .06 | .626 | -.05 | .11 | .679 |
| 140 | Very few headaches. * | -.03 | .06 | .572 | -.03 | .06 | .588 |
| 156 | Hands and feet warm. * | -.03 | .06 | .679 | .12 | .07 | .076 |
| 163 | Not notice ears ringing. * | -.03 | .06 | .590 | -.05 | .07 | .514 |
| 167 | Troubled by discomfort in stomach. | -.03 | .06 | .642 | -.05 | .07 | .452 |
| 173 | Parts of body fell like they are burning. | -.04 | .06 | .556 | .09 | .07 | .189 |
| 176 | Seldom have dizzy spells. * | -.03 | .06 | .656 | -.03 | .07 | .620 |
| 187 | Have been paralyzed or unusual weakness. * | -.05 | .06 | .445 | .15 | .06 | .012 |
| 189 | Feel tight band around head. | -.03 | .06 | .608 | .00 | .06 | .986 |
| 192 | Stomach trouble. | -.03 | .06 | .687 | .12 | .07 | .056 |
| 207 | Never bother by pain over heart. * | -.03 | .06 | .584 | .02 | .07 | .723 |
| 227 | Fullness in head or nose. | -.04 | .06 | .558 | .07 | .08 | .432 |

Table 6
Differential Item Functioning: RC2

| Item | | RC2 on Age | | | Item on Age | | |
|------|--|------------|------------|-------------|-------------|------------|-------------|
| | | Estimate | S.E. | <i>p</i> | Estimate | S.E. | <i>p</i> |
| 1 | Enjoy play and recreation. * | .14 | .07 | .032 | .00 | .07 | .982 |
| 2 | Sociable person. * | .12 | .08 | .111 | .00 | .07 | .974 |
| 56 | Goals in life are in reach. * | .13 | .08 | .080 | -.05 | .07 | .527 |
| 84 | Make up mind with ease. * | .12 | .07 | .102 | .03 | .07 | .676 |
| 137 | Excitement pulls me out when I am low. * | .11 | .07 | .135 | .09 | .07 | .191 |
| 157 | Seems to make friends. * | .10 | .08 | .191 | .07 | .07 | .278 |
| 169 | Expect to succeed. * | .15 | .07 | .047 | -.16 | .08 | .030 |
| 175 | Get along with people. * | .14 | .08 | .073 | -.05 | .07 | .433 |
| 220 | Enjoy social gatherings. * | .14 | .08 | .071 | -.05 | .06 | .433 |
| 237 | No close friends. | .10 | .08 | .167 | -.18 | .08 | .025 |

Note. Significant *p*-values are bolded. RC2 = Low Positive Emotions; Estimate = Standardized Estimate; S.E. = Standard Error; *p* = probability. All items contribute to the total score. If they are endorsed true, with the exception of those marked with an *, which contribute to the total score when endorsed false.

Table 7*Differential Item Functioning: RC3*

| Item | | RC3 on Age | | | Item on Age | | |
|------|---|------------|------|----------|-------------|------------|-------------|
| | | Estimate | S.E. | <i>p</i> | Estimate | S.E. | <i>p</i> |
| 24 | Most are honest because of fear of getting caught. | .12 | .07 | .068 | -.21 | .07 | .002 |
| 31 | People blame others for their mistakes. | .12 | .07 | .068 | -.21 | .07 | .002 |
| 90 | Most lie to get ahead. | .08 | .07 | .250 | .03 | .08 | .746 |
| 100 | Safer to trust nobody. | .10 | .07 | .154 | -.06 | .07 | .387 |
| 105 | Make friends because friends are useful. | .08 | .07 | .261 | .02 | .07 | .757 |
| 131 | People demand respect for their rights than for others. | .09 | .07 | .160 | -.08 | .08 | .336 |
| 152 | Takes argument to convince others of the truth. | .05 | .07 | .500 | .15 | .07 | .026 |
| 191 | People dislike putting themselves out to help others. | .09 | .07 | .169 | -.04 | .07 | .561 |
| 196 | Anyone would tell a lie to keep out of trouble. | .06 | .07 | .407 | .16 | .08 | .035 |

Note. Significant *p*-values are bolded. RC3 = Cynicism; Estimate = Standardized Estimate; S.E. = Standard Error; *p* = probability. All items contribute to the total score. If they are endorsed true, with the exception of those marked with an *, which contribute to the total score when endorsed false.

Table 8*Differential Item Functioning: RC4*

| Item | | RC4 on Age | | | Item on Age | | |
|------|---|------------|------|----------|-------------|------------|-------------|
| | | Estimate | S.E. | <i>p</i> | Estimate | S.E. | <i>p</i> |
| 14 | Have stolen things. | -.01 | .06 | .863 | -.01 | .07 | .857 |
| 19 | Parents do not like friends. | -.01 | .06 | .935 | -.06 | .06 | .342 |
| 33 | Never in trouble with the law. * | -.01 | .06 | .872 | -.12 | .09 | .167 |
| 40 | Have not gone to school when should have. | -.01 | .06 | .895 | .09 | .09 | .330 |
| 43 | Problem with alcohols and drugs. | -.01 | .06 | .927 | -.07 | .06 | .242 |
| 64 | Friends talk me into things that are wrong. | -.02 | .06 | .720 | .11 | .07 | .102 |
| 72 | People tell me I have a problem with drinking. | -.02 | .06 | .720 | .11 | .07 | .102 |
| 88 | Have been suspended from school. | .00 | .06 | .994 | -.15 | .08 | .076 |
| 106 | Get into fights when drinking. | -.02 | .06 | .718 | .15 | .07 | .035 |
| 110 | School grades in classroom behaviors are bad. | -.02 | .06 | .752 | .09 | .07 | .202 |
| 111 | Friends often in trouble. | -.02 | .06 | .704 | .10 | .06 | .066 |
| 118 | Spent nights away from home without parent knowledge. | -.03 | .06 | .637 | .14 | .05 | .008 |
| 127 | Sent to principal for bad behavior. | -.01 | .06 | .839 | .01 | .06 | .850 |
| 144 | Better if all laws were thrown away. | -.02 | .06 | .716 | .17 | .08 | .033 |
| 146 | Parents object to who I hang with. | -.01 | .06 | .845 | .01 | .07 | .890 |
| 148 | Impossible from stealing or shoplifting. | .00 | .06 | .977 | -.08 | .06 | .199 |
| 160 | Have done bad things because friends. | -.02 | .06 | .747 | .13 | .07 | .065 |
| 166 | Used alcohol excessively. | -.01 | .06 | .831 | .02 | .07 | .743 |
| 235 | Enjoy marijuana. | .01 | .06 | .877 | -.19 | .06 | .004 |
| 238 | Never run away from home. * | .01 | .06 | .935 | -.16 | .06 | .009 |

Note. Significant *p*-values are bolded. RC4 = Antisocial Behavior; Estimate = Standardized Estimate; S.E. = Standard Error; *p* = probability. All items contribute to the total score. If they are endorsed true, with the exception of those marked with an *, which contribute to the total score when endorsed false.

Table 9*Differential Item Functioning: RC6*

| Item | | RC6 on Age | | | Item on Age | | |
|------|--|------------|------|----------|-------------|------|----------|
| | | Estimate | S.E. | <i>p</i> | Estimate | S.E. | <i>p</i> |
| 26 | People are not kind to me. | .01 | .06 | .877 | -.07 | .07 | .299 |
| 37 | Someone trying to punish me. | .02 | .06 | .787 | -.04 | .06 | .513 |
| 45 | People say vulgar things about me. | .01 | .06 | .877 | -.07 | .07 | .299 |
| 57 | Teachers have it in for me. | -.02 | .06 | .791 | .12 | .08 | .116 |
| 76 | No enemies who wish harm. * | -.01 | .06 | .816 | -.08 | .07 | .258 |
| 92 | Punished without cause. | -.01 | .06 | .882 | .04 | .07 | .597 |
| 134 | Plotted against. | -.02 | .06 | .758 | .08 | .07 | .229 |
| 164 | If people had not, I would be more successful. | .01 | .06 | .855 | -.07 | .07 | .318 |
| 199 | I am being talked about. | .01 | .06 | .928 | -.05 | .07 | .505 |

Note. Significant *p*-values are bolded. RC6 = Ideas of Persecution; Estimate = Standardized Estimate; S.E. = Standard Error; *p* = probability. All items contribute to the total score. If they are endorsed true, with the exception of those marked with an *, which contribute to the total score when endorsed false.

Table 10*Differential Item Functioning: RC7*

| Item | | RC7 on Age | | | Item on Age | | |
|------|---|------------|------|----------|-------------|------------|-------------|
| | | Estimate | S.E. | <i>p</i> | Estimate | S.E. | <i>p</i> |
| 6 | Fight against bashfulness. | .01 | .06 | .925 | .09 | .06 | .182 |
| 23 | Easily embarrassed. | .00 | .06 | .986 | -.09 | .06 | .135 |
| 32 | Criticism hurts. | -.01 | .06 | .894 | -.05 | .06 | .438 |
| 55 | Something happens to frighten me. | -.02 | .06 | .713 | .07 | .08 | .375 |
| 71 | Feels that something dreadful is going to happen. | -.02 | .06 | .724 | .04 | .06 | .495 |
| 115 | I do things that I regret. | -.01 | .06 | .842 | -.02 | .06 | .736 |
| 133 | Often sorry because I am irritable. | -.02 | .06 | .738 | .04 | .06 | .548 |
| 141 | Touchy on subjects that I can't talk about them. | -.02 | .06 | .702 | .07 | .06 | .294 |
| 153 | Feel anxiety about something or someone. | .01 | .06 | .865 | -.15 | .06 | .007 |
| 161 | Easily become impatient. | -.02 | .06 | .772 | .03 | .07 | .705 |
| 216 | Worrying about things said that hurt others feelings. | -.01 | .06 | .823 | -.01 | .06 | .904 |

Note. Significant *p*-values are bolded. RC7 = Dysfunctional Negative Emotions; Estimate = Standardized Estimate; S.E. = Standard Error; *p* = probability. All items contribute to the total score. If they are endorsed true, with the exception of those marked with an *, which contribute to the total score when endorsed false.

Table 11*Differential Item Functioning: RC8*

| Item | | RC8 on Age | | | Item on Age | | |
|------|--|------------|------|----------|-------------|------|----------|
| | | Estimate | S.E. | <i>p</i> | Estimate | S.E. | <i>p</i> |
| 27 | Soul leaves body. | .03 | .04 | .464 | -.09 | .09 | .339 |
| 49 | Attacks in which can't control body but knows what's going on. | .04 | .08 | .608 | .00 | .07 | .999 |
| 68 | Something was making me do things through hypnotizing me. | .05 | .07 | .520 | -.06 | .10 | .549 |
| 108 | Bothered by hearing strange things. | .00 | .08 | .966 | .15 | .08 | .075 |
| 142 | Someone controlling thoughts. | .04 | .07 | .629 | .02 | .11 | .877 |
| 204 | Someone tries to influence my mind. | .04 | .08 | .634 | .01 | .08 | .929 |
| 222 | Hear voices without knowing their location. | .06 | .08 | .430 | -.07 | .08 | .397 |

Note. Significant *p*-values are bolded. RC8 = Aberrant Experiences; Estimate = Standardized Estimate; S.E. = Standard Error; *p* = probability. All items contribute to the total score. If they are endorsed true, with the exception of those marked with an *, which contribute to the total score when endorsed false. The model for RC8 was recalculated without Item 168, as it was a Heywood case. Item 168 = seeing things that others could not.

Table 12*Differential Item Functioning: RC9*

| Item | | RC9 on Age | | | Item on Age | | |
|------|--|------------|------|-------------|-------------|------|----------|
| | | Estimate | S.E. | <i>p</i> | Estimate | S.E. | <i>p</i> |
| 58 | Stood in the way of others because of principle. | .07 | .07 | .337 | -.03 | .07 | .615 |
| 61 | Sleep did not seem necessary for days. | .08 | .07 | .256 | -.09 | .07 | .196 |
| 86 | Without reason, I feel happy. | .04 | .07 | .579 | .07 | .06 | .257 |
| 94 | When bored, like to stir some excitement. | .04 | .07 | .612 | .10 | .07 | .125 |
| 124 | Becomes very excited. | .05 | .07 | .481 | .03 | .07 | .635 |
| 177 | Fits of uncontrollable laughing and crying. | .08 | .07 | .271 | -.08 | .07 | .239 |
| 182 | Never done anything dangerous for thrill* | .06 | .07 | .414 | -.04 | .07 | .621 |
| 236 | So excited, find it hard to sleep. | .07 | .08 | .342 | -.03 | .07 | .625 |

Note. Significant *p*-values are bolded. RC9 = Hypomanic Activation; Estimate = Standardized Estimate; S.E. = Standard Error; *p* = probability. All items contribute to the total score. If they are endorsed true, with the exception of those marked with an *, which contribute to the total score when endorsed false.

Table 13
Moderation Analysis: RCd, Age, and Criterion Measures

| Initial Models | | | | | | | | | | Moderation Model | | | | | | | |
|------------------|--------|----------------|-------|-------------|-----------------------------|---------|------|---------------------------|------|------------------|----------------|-------|-------------|-----------------------------|---------|------|-----------------|
| Outcome | R | R ² | F | df1, df2 | Unstandardized Coefficients | | | Standardized Coefficients | | Model Fit | | | | Unstandardized Coefficients | | | |
| | | | | | Constant | RCd | Age | RCd | Age | R | R ² | F | df1, df2 | Constant | RCd | Age | RCd x Age |
| BASC- 3DEP | .72*** | .52 | 92.84 | 2, 172 | -1.04 | 1.06*** | .25 | .72*** | .04 | .72** * | .52 | 62.22 | 3, 171 | 10.38*** | 1.05*** | .26 | .07 |
| BASC-3 LOC | .56*** | .31 | 39.40 | 2, 172 | 6.42 | .51*** | -.20 | .56*** | -.05 | .56** * | .32 | 26.39 | 3, 171 | 7.09*** | .51*** | -.20 | -.04 |
| BASC-3 SE | .67*** | .45 | 71.27 | 2, 172 | 18.50*** | 0.65*** | -.11 | .67*** | -.03 | .67** * | .45 | 47.45 | 3, 171 | 12.23*** | -.64*** | -.12 | -.03 |
| BASC-3 SI | .70*** | .49 | 81.35 | 2, 172 | 6.45 | .86*** | .00 | .70*** | .00 | .70** * | .49 | 54.1 | 3, 171 | 12.50*** | .86*** | -.01 | -.03 |
| MINI DEP Cur. | .62*** | .39 | 34.18 | 2, 108 | -.19 | .39*** | -.00 | .62*** | .00 | .62** * | .39 | 22.58 | 3, 107 | 2.95*** | .38*** | -.00 | .00 |
| MINI DEP Past | .48*** | .23 | 18.04 | 2, 119 | .74 | .29*** | .13 | .48*** | .04 | .49** * | .24 | 12.47 | 3, 118 | 4.98*** | .30*** | .13 | -.05 |
| MINI SUC | .46*** | .21 | 12.25 | 2, 91 | -.29 | .37*** | -.03 | .46*** | -.01 | .47** * | .22 | 8.62 | 3, 90 | 2.28*** | .38*** | -.03 | -.08 |

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. BASC-3 DEP = BASC-3 Depression Scale; BASC-3 SE = BASC-3 Self-Esteem Scale; BASC-3 SI = BASC-3 Sense of Inadequacy Scale; MINI DEP Cur.= MINI-KID Depression Current; MINI DEP Past = MINI-KID Depression Past; MINI SUC = MINI-KID Suicidality and Suicide Behavior Disorder.

Table 14
Moderation Analysis: RC1, Age, and Criterion Measures

| Initial Models | | | | | | | | | | Moderation Model | | | | | | | |
|----------------|--------|----------------|-------|----------------------------|-----------------------------|--------|------|---------------------------|------|------------------|----------------|-------|----------------------------|-----------------------------|--------|------|-----------------|
| | | | | | Unstandardized Coefficients | | | Standardized Coefficients | | Model Fit | | | | Unstandardized Coefficients | | | |
| Outcome | R | R ² | F | <i>df1</i> , <i>df2</i> | Constant | RC1 | Age | RC1 | Age | R | R ² | F | <i>df1</i> , <i>df2</i> | Constant | RC1 | Age | RC1 x Age |
| BASC-3 SOM | .67*** | .45 | 69.85 | 2, 175 | .35 | .50*** | -.03 | .67*** | -.01 | .67*** | .45 | 46.47 | 3, 171 | 2.98*** | .50*** | -.03 | -.02 |

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. BASC-3 SOM = BACS-3 Somatization Scale.

Table 15*Moderation Analysis: RC2, Age, and Criterion Measures*

| Initial Models | | | | | Moderation Model | | | | | | | | | | | | |
|------------------|--------|----------------|------|-------------|-----------------------------|---------|-----|---------------------------|-----|-----------|----------------|------|-------------|-----------------------------|---------|-----|-----------------|
| Outcome | R | R ² | F | df1, df2 | Unstandardized Coefficients | | | Standardized Coefficients | | Model Fit | | | | Unstandardized Coefficients | | | |
| | | | | | Constant | RC2 | Age | RC2 | Age | R | R ² | F | df1, df2 | Constant | RC2 | Age | RC2 x Age |
| BASC- 3 DEP | .30*** | .09 | 8.34 | 2, 172 | -2.42 | 1.23*** | .65 | .29*** | .10 | .32** | .10 | 6.43 | 3, 171 | 10.47*** | 1.23*** | .79 | .37 |
| MINI DEP Cur. | .35*** | .12 | 7.55 | 2, 108 | -2.15 | 0.59*** | .23 | .35*** | .08 | .38*** | .15 | 6.19 | 3, 107 | 2.90*** | .52** | .20 | -.22 |
| MINI DEP Past | .30** | .09 | 6.03 | 2, 119 | -1.05 | 0.49*** | .31 | .49*** | .11 | .32** | .10 | 4.49 | 3, 118 | 4.93*** | .44** | .29 | -.14 |

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. BASC-3 DEP = BACS-3 Depression Scale; MINI DEP Cur.= MINI-KID Depression Current; MINI DEP Past = MINI-KID Depression Past.

Table 16
Moderation Analysis: RC3, Age, and Criterion Measures

| Initial Models | | | | | | | | | | Moderation Model | | | | | | | |
|----------------|--------|----------------|-------|-------------|-----------------------------|--------|------|---------------------------|------|------------------|----------------|------|-------------|-----------------------------|-------|------|--------------|
| Outcome | R | R ² | F | df1, df2 | Unstandardized Coefficients | | | Standardized Coefficients | | Model Fit | | | | Unstandardized Coefficients | | | |
| | | | | | Constant | RC3 | Age | RC3 | Age | R | R ² | F | df1, df2 | Constant | RC3 | Age | RC3 x Age |
| BASC-3 IR | .22* | .05 | 4.43 | 2, 172 | 23.63*** | .498** | -.24 | -.22** | -.06 | .23* | .05 | 3.17 | 3, 171 | 17.14*** | .49** | -.23 | -.13 |
| BASC-3 AT | .33*** | .11 | 10.51 | 2, 171 | 14.12* | .83*** | -.58 | .30*** | -.12 | .25* | .06 | 3.86 | 3, 171 | 6.26*** | .64** | -.21 | .02 |
| BASC-3 SR | .12 | .01 | 1.17 | 2, 172 | 11.13* | -.18 | .29 | -.09 | .07 | .14 | .02 | 1.09 | 3, 171 | 14.47*** | -.17 | .27 | .15 |

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. BASC-3 IR = BASC-3 Interpersonal Relations Scales; BASC-3 AT = BASC-3 Attitudes Towards Teachers; BASC-3 SR = BASC-3 Self-Reliance Scale.

Table 17*Moderation Analysis: RC4, Age, and Criterion Measures*

| Initial Models | | | | | | | | | | Moderation Model | | | | | | | |
|----------------|--------|----------------|-------|-------------|-----------------------------|--------|-------|---------------------------|------|------------------|----------------|-------|-------------|-----------------------------|--------|-------|-----------------|
| Outcome | R | R ² | F | df1, df2 | Unstandardized Coefficients | | | Standardized Coefficients | | Model Fit | | | | Unstandardized Coefficients | | | |
| | | | | | Constant | RC4 | Age | RC4 | Age | R | R ² | F | df1, df2 | Constant | RC4 | Age | RC4 x Age |
| BASC-3 AC | .42*** | .17 | 18.00 | 2, 172 | 13.36 | .55*** | -.43 | .40*** | -.09 | .42*** | .17 | 11.95 | 3, 171 | 11.61*** | .55*** | -.43 | -.02 |
| MINI CD | .25* | .06 | 3.87 | 2, 118 | -296.65 | -6.47 | 25.74 | -.16 | .17 | .29* | .09 | 3.65 | 3, 117 | 34.45* | -7.19* | 25.44 | -5.53 |
| MINI ODD | .38*** | .14 | 9.08 | 2, 108 | 6.10 | .26*** | -.25 | .36*** | -.09 | .38** | .14 | 6.04 | 3, 107 | 4.63*** | .26** | -.26 | -.02 |
| MINI AUD | .77*** | .56 | 16.66 | 2, 26 | -10.47 | .52*** | .55 | .79*** | .21 | .77*** | .59 | 11.89 | 3, 25 | 4.65*** | .47*** | .58 | .11 |
| MINI SUD | .45*** | .20 | 14.03 | 2, 111 | -3.75 | .37*** | .24 | .45*** | .08 | .46*** | .21 | 9.89 | 3, 110 | 3.19*** | .35*** | .22 | -.07 |

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. BASC-3 AC = BASC-3 Anger Control Scale; BASC-3 AA = BASC-3 Alcohol Abuse Scale; BASC-3 CP = BASC-3 Conduct Problems Scale; MINI CD = MINI-KID Conduct Disorder; MINI ODD = MINI-KID Oppositional Defiance Disorder; MINI SUD = MINI-KID Other Substance Use Disorder.

Table 18
Moderation Analysis: RC6, Age, and Criterion Measures

| Initial Models | | | | | | | | | | Moderation Model | | | | | | | |
|----------------|--------|----------------|-------|-------------|-----------------------------|---------|------|---------------------------|------|------------------|----------------|-------|-------------|-----------------------------|---------|------|-----------------|
| Outcome | R | R ² | F | df1, df2 | Unstandardized Coefficients | | | Standardized Coefficients | | Model Fit | | | | Unstandardized Coefficients | | | |
| | | | | | Constant | RC6 | Age | RC6 | Age | R | R ² | F | df1, df2 | Constant | RC6 | Age | RC6 x Age |
| BASC-3 AS | .43*** | .18 | 19.13 | 2, 172 | 6.80 | 1.10*** | -.26 | .42*** | -.05 | .43*** | .18 | 12.71 | 3, 171 | 6.26*** | 1.07*** | -.26 | .05 |
| BASC-3 AT | .33 | .11 | 10.70 | 2, 171 | 17.80** | .81*** | -.67 | .30*** | -.13 | .33** | .11 | 7.10 | 3, 170 | 10.15*** | .82*** | -.67 | -.03 |

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. BASC-3 AS = BASC-3 Atypicality Scale; BASC-3 AT = BASC-3 Attitudes Towards Teachers Scale.

Table 19*Moderation Analysis: RC7, Age, and Criterion Measures*

| Initial Models | | | | | | | | | | Moderation Model | | | | | | | |
|----------------|--------|----------------|-------|-------------|-----------------------------|---------|--------|---------------------------|-------|------------------|----------------|-------|-------------|-----------------------------|---------|------|--------------|
| Outcome | R | R ² | F | df1, df2 | Unstandardized Coefficients | | | Standardized Coefficients | | Model Fit | | | | Unstandardized Coefficients | | | |
| | | | | | Constant | RC7 | Age | RC7 | Age | R | R ² | F | df1, df2 | Constant | RC7 | Age | RC7 x Age |
| BASC-3 AX | .64*** | .41 | 60.05 | 2, 172 | -10.86 | 1.78*** | 1.09** | .62*** | .16** | .64*** | .41 | 39.98 | 3, 171 | 14.58*** | 1.77*** | 1.09 | -.08 |
| MINI AGOR | .38*** | .15 | 9.66 | 2, 113 | -2.40 | .24*** | .16 | .37*** | .09 | .41*** | .17 | 7.51 | 3, 112 | 1.18*** | .24*** | .13 | .08 |
| MINI GAD | .52*** | .27 | 21.26 | 2, 114 | -9.35 | .68*** | .58* | .50*** | .16* | .53*** | .28 | 14.76 | 3, 113 | 3.05*** | .69*** | .55 | .12 |
| MINI PAN | .33** | .11 | 6.87 | 2, 115 | -.57 | .53*** | .02 | .33*** | .00 | .34** | .12 | 4.98 | 3, 114 | 2.38*** | .54** | -.03 | .14 |
| MINI SEP | .31** | .09 | 5.81 | 2, 112 | -.89 | .23*** | .08 | .30*** | .04 | .31** | .10 | 3.96 | 3, 111 | 1.43*** | .23*** | .07 | .03 |
| MINI OCD | .24* | .06 | 3.50 | 2, 112 | -.89 | .11* | .06 | .24* | .05 | .24 | .06 | 2.32 | 3, 111 | .65*** | .11* | .07 | -.00 |
| MINI PTSD | .46*** | .21 | 14.99 | 2, 110 | 6.67 | .47*** | -.35 | .47*** | -.07 | .49*** | .24 | 11.47 | 3, 109 | 4.91*** | .88*** | -.14 | .04 |

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. BASC-3 AX = BASC-3 Anxiety Scale; MINI AGOR = MINI-KID Agoraphobia; MINI GAD = MINI-KID Generalized Anxiety Disorder; MINI PAN = MINI-KID Panic Disorder; MINI SEP = MINI-KID Separation Anxiety Disorder; MINI OCD = MINI-KID Obsessive-Compulsive Disorder; MINI PTSD = MINI-KID Post-Traumatic Stress Disorder.

Table 20*Moderation Analysis: RC8, Age, and Criterion Measures*

| Initial Models | | | | | | | | | | Moderation Model | | | | | | | |
|----------------------|--------|----------------|-------|-------------|-----------------------------|---------|------|---------------------------|------|------------------|----------------|-------|-------------|-----------------------------|---------|------|-----------------|
| Outcome | R | R ² | F | df1, df2 | Unstandardized Coefficients | | | Standardized Coefficients | | Model Fit | | | | Unstandardized Coefficients | | | |
| | | | | | Constant | RC8 | Age | RC8 | Age | R | R ² | F | df1, df2 | Constant | RC8 | Age | RC8 x Age |
| BASC- 3 AS | .66*** | .44 | 67.15 | 2, 172 | 3.96 | 2.79*** | -.04 | .66*** | -.01 | .66*** | .44 | 44.52 | 3, 171 | 6.26*** | 2.80*** | -.04 | .04 |
| MINI PSYC Past | .50*** | .25 | 16.96 | 2, 103 | -.38 | .54*** | .04 | .50*** | .03 | .50*** | .25 | 11.51 | 3, 102 | .97*** | .54*** | .05 | -.07 |
| MINI PSYC Cur. | .43* | .18 | 4.71 | 2, 44 | -.81 | .37** | .09 | .42** | .07 | .43* | .19 | 3.07 | 3, 41 | 1.25*** | .37*** | .08 | -.01 |

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. BASC-3 AS = BACS-3 Atypicality Scale; MINI PSYC Past = MINI-KID Psychosis Past; MINI PSYC Cur. = MINI-KID Psychosis Current.

Table 21*Moderation Analysis: RC9, Age, and Criterion Measures*

| Outcome | Initial Models | | | | | | | | | Moderation Model | | | | | | | |
|---------------------|----------------|----------------|-------|-------------|-----------------------------|---------|------|---------------------------|------|------------------|----------------|-------|-------------|-----------------------------|---------|------|-----------------|
| | R | R ² | F | df1, df2 | Unstandardized Coefficients | | | Standardized Coefficients | | Model Fit | | | | Unstandardized Coefficients | | | |
| | | | | | Constant | RC9 | Age | RC9 | Age | R | R ² | F | df1, df2 | Constant | RC9 | Age | RC9 x Age |
| BASC-3 MAN | .43*** | .18 | 19.03 | 2, 172 | 10.80 | 1.40*** | -.26 | .42*** | -.05 | .43*** | .18 | 12.62 | 3, 171 | 11.95*** | 1.40*** | -.26 | -.02 |
| BASC-3 SS | .41*** | .17 | 16.93 | 2, 172 | 11.86* | 1.15*** | -.32 | .40*** | -.07 | .41*** | .17 | 11.27 | 3, 171 | 11.10*** | 1.15*** | -.33 | -.06 |
| BASC-3 HYP | .37*** | .14 | 13.98 | 2, 172 | 14.61 | .89*** | -.55 | .35*** | -.13 | .38*** | .14 | 9.38 | 3, 171 | 9.47*** | .89*** | -.56 | -.08 |
| MINI MAN Past | .32** | .10 | 6.32 | 2, 114 | .82 | .49*** | .01 | .32 | .00 | .32** | .10 | 4.20 | 3, 113 | 2.80*** | .49*** | .01 | .03 |
| MINI MAN Cur. | .17 | .03 | 1.16 | 2, 83 | 1.04 | .15 | -.06 | .16 | -.04 | .24 | .06 | 1.69 | 3, 82 | .75*** | .14 | -.05 | .14 |

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. BASC-3 MAN = BACS-3 Mania Scale; BASC-3 SS = BACS-3 Sensation Seeking Scale; BASC-3 HYP = BACS-3 Hyperactivity Scale; MINI MAN Past = MINI-KID Mania Past; MINI MAN Cur. = MINI-KID Mania Current.

Appendix A:
Individual Scale Standardized Parameter Estimates

Table A1
Standardized Parameter Estimates: RCD

| Items | Abbreviated Item Descriptions | Estimate | S.E. | <i>p</i> |
|-------|--|----------|------|----------|
| 8 | Hard to keep mind on task. | .53 | .06 | .001 |
| 28 | Could not overcome difficulties. | .69 | .05 | .001 |
| 34 | Even when with people, feel lonely. | .77 | .04 | .001 |
| 60 | Feel helpless when making important decisions. | .72 | .05 | .001 |
| 69 | Stand up for what is right. | .69 | .05 | .001 |
| 73 | Lacking self-confidence. | .76 | .04 | .001 |
| 74 | Don't care what happens. | .69 | .05 | .001 |
| 79 | Feels useless. | .91 | .02 | .001 |
| 82 | Not happy with self. | .74 | .04 | .001 |
| 102 | Difficulty starting things. | .59 | .06 | .001 |
| 128 | About to go to pieces. | .73 | .04 | .001 |
| 138 | Life is a strain. | .80 | .04 | .001 |
| 162 | Not happy with life, it's too late. | .59 | .06 | .001 |
| 170 | Something wrong with mind. | .77 | .05 | .001 |
| 183 | Wish to be happy. | .85 | .03 | .001 |
| 190 | Happy most of the time. * | -.78 | .04 | .001 |
| 230 | Everything is fine, I don't care. | .82 | .04 | .001 |
| 234 | I am no good. | .85 | .03 | .001 |

Notes. RCD = Demoralization; Estimate = Standardized Estimate; S.E. = Standard Error; *p* = probability. All items contribute to the total score. If they are endorsed true, with the exception of those marked with an *, which contribute to the total score when endorsed false.

Table A2*Standardized Parameter Estimates: RC1*

| Items | Abbreviated Item Description | Estimate | S.E. | <i>p</i> |
|-------|---|----------|------|----------|
| 10 | No difficulties keeping balance. * | -.39 | .08 | .001 |
| 20 | Lump in my throat. | .60 | .08 | .001 |
| 21 | Periods when couldn't do things because couldn't get going. | .58 | .06 | .001 |
| 22 | Head seems to hurt. | .75 | .05 | .001 |
| 23 | Easily embarrassed. | .58 | .06 | .001 |
| 25 | Few or no pains. * | -.59 | .06 | .001 |
| 59 | Bothered by an upset stomach. | .79 | .05 | .001 |
| 63 | Top of head feels tender. | .53 | .07 | .001 |
| 66 | Hardly feel pain the neck. * | -.51 | .07 | .001 |
| 78 | Hardly no heart pounding and short of breath. * | -.09 | .08 | .255 |
| 85 | No trouble with muscles twitching. * | -.45 | .07 | .001 |
| 95 | Attacks of nausea. | .70 | .10 | .001 |
| 140 | Very few headaches. * | -.61 | .06 | .001 |
| 156 | Hands and feet warm. * | -.35 | .08 | .001 |
| 163 | Not notice ears ringing. * | -.24 | .08 | .001 |
| 167 | Troubled by discomfort in stomach. | .39 | .07 | .001 |
| 173 | Parts of body fell like they are burning. | .34 | .07 | .001 |
| 176 | Seldom have dizzy spells. * | .82 | .04 | .001 |
| 187 | Have been paralyzed or unusual weakness. * | .69 | .05 | .001 |
| 189 | Feel tight band around head. | -.57 | .06 | .001 |
| 192 | Stomach trouble. | -.38 | .08 | .001 |
| 207 | Never bother by pain over heart. * | .78 | .05 | .001 |
| 227 | Fullness in head or nose. | .79 | .05 | .001 |
| 232 | Hands not clumsy or awkward. * | -.48 | .06 | .001 |

Notes. RC1 = Somatic Complaints; Estimate = Standardized Estimate; S.E. = Standard Error; *p* = probability. All items contribute to the total score. If they are endorsed true, with the exception of those marked with an *, which contribute to the total score when endorsed false. Because Item 95 of RC1 was a Heywood case, Item 95's unstandardized loading was set to 1.



Table A3*Standardized Parameter Estimates: RC2*

| Items | Abbreviated Item Description | Estimate | S.E. | <i>p</i> |
|-------|--|----------|------|----------|
| 1 | Enjoy play and recreation. * | .30 | .09 | .001 |
| 2 | Sociable person. * | .75 | .06 | .001 |
| 56 | Goals in life are in reach. * | .44 | .08 | .001 |
| 84 | Make up mind with ease. * | .24 | .09 | .006 |
| 137 | Excitement pulls me out when I am low. * | .39 | .08 | .001 |
| 157 | Seems to make friends. * | .78 | .06 | .001 |
| 169 | Expect to succeed. * | .44 | .09 | .001 |
| 175 | Get along with people. * | .62 | .07 | .001 |
| 220 | Enjoy social gatherings. * | .63 | .07 | .001 |
| 237 | No close friends. | -.36 | .10 | .001 |

Notes. RC2 = Low Positive Emotions; Estimate = Standardized Estimate; S.E. = Standard Error; *p* = probability. All items contribute to the total score. If they are endorsed true, with the exception of those marked with an *, which contribute to the total score when endorsed false.

Table A4*Standardized Parameter Estimates: RC3*

| Items | Abbreviated Item Description | Estimate | S.E. | <i>p</i> |
|-------|---|----------|------|----------|
| 24 | Most are honest because of fear of getting caught. | .43 | .08 | .001 |
| 31 | People blame others for their mistakes. | .42 | .08 | .001 |
| 90 | Most lie to get ahead. | .71 | .06 | .001 |
| 100 | Safer to trust nobody. | .53 | .07 | .001 |
| 105 | Make friends because friends are useful. | .60 | .07 | .001 |
| 131 | People demand respect for their rights than for others. | .43 | .08 | .001 |
| 152 | Takes argument to convince others of the truth. | .63 | .07 | .001 |
| 191 | People dislike putting themselves out to help others. | .61 | .07 | .001 |
| 196 | Anyone would tell a lie to keep out of trouble. | .54 | .07 | .001 |

Notes: RC3 = Cynicism; Estimate = Standardized Estimate; S.E. = Standard Error; *p* = probability. All items contribute to the total score. If they are endorsed true, with the exception of those marked with an *, which contribute to the total score when endorsed false.

Table A5*Standardized Parameter Estimates: RC4*

| Items | Abbreviated Item Description | Estimate | S.E. | <i>p</i> |
|-------|---|----------|------|----------|
| 14 | Have stolen things. | .41 | .07 | .001 |
| 19 | Parents do not like friends. | .64 | .05 | .001 |
| 33 | Never in trouble with the law. * | .13 | .10 | .184 |
| 40 | Have not gone to school when should have. | -.43 | .08 | .001 |
| 43 | Problem with alcohols and drugs. | .48 | .61 | .001 |
| 64 | Friends talk me into things that are wrong. | .71 | .05 | .001 |
| 72 | People tell me I have a problem with drinking. | .68 | .05 | .001 |
| 88 | Have been suspended from school. | .87 | .04 | .001 |
| 106 | Get into fights when drinking. | .58 | .07 | .001 |
| 110 | School grades in classroom behaviors are bad. | .72 | .05 | .001 |
| 111 | Friends often in trouble. | .71 | .05 | .001 |
| 118 | Spent nights away from home without parent knowledge. | .73 | .04 | .001 |
| 127 | Sent to principal for bad behavior. | .55 | .06 | .001 |
| 144 | Better if all laws were thrown away. | .59 | .07 | .001 |
| 146 | Parents object to who I hang with. | .47 | .07 | .001 |
| 148 | Impossible from stealing or shoplifting. | .71 | .04 | .001 |
| 160 | Have done bad things because friends. | .47 | .07 | .001 |
| 166 | Used alcohol excessively. | .55 | .06 | .001 |
| 235 | Enjoy marijuana. | .77 | .04 | .001 |
| 238 | Never run away from home. * | .61 | .05 | .001 |

Notes: RC4 = Antisocial Behavior; Estimate = Standardized Estimate; S.E. = Standard Error; *p* = probability. All items contribute to the total score. If they are endorsed true, with the exception of those marked with an *, which contribute to the total score when endorsed false.

Table A6*Standardized Parameter Estimates: RC6*

| Items | Abbreviated Item Description | Estimate | S.E. | <i>p</i> |
|-------|--|----------|------|----------|
| 26 | People are not kind to me. | .62 | .06 | .001 |
| 37 | Someone trying to punish me. | .77 | .05 | .001 |
| 45 | People say vulgar things about me. | .58 | .06 | .001 |
| 57 | Teachers have it in for me. | .45 | .08 | .001 |
| 76 | No enemies who wish harm. * | -.45 | .07 | .001 |
| 92 | Punished without cause. | .50 | .07 | .001 |
| 134 | Plotted against. | .76 | .06 | .001 |
| 164 | If people had not, I would be more successful. | .72 | .06 | .001 |
| 199 | I am being talked about. | .63 | .06 | .001 |

Notes: RC6 = Ideas of Persecution; Estimate = Standardized Estimate; S.E. = Standard Error; *p* = probability. All items contribute to the total score. If they are endorsed true, with the exception of those marked with an *, which contribute to the total score when endorsed false.

Table A7*Standardized Parameter Estimates: RC7*

| Items | Abbreviated Item Description | Estimate | S.E. | <i>p</i> |
|-------|---|----------|------|----------|
| 6 | Fight against bashfulness. | .53 | .07 | .001 |
| 23 | Easily embarrassed. | .73 | .05 | .001 |
| 32 | Criticism hurts. | .65 | .06 | .001 |
| 55 | Something happens to frighten me. | .73 | .07 | .001 |
| 71 | Feels that something dreadful is going to happen. | .79 | .05 | .001 |
| 115 | I do things that I regret. | .63 | .06 | .001 |
| 133 | Often sorry because I am irritable. | .72 | .05 | .001 |
| 141 | Touchy on subjects that I can't talk about them. | .59 | .06 | .001 |
| 153 | Feel anxiety about something or someone. | .75 | .05 | .001 |
| 161 | Easily become impatient. | .60 | .06 | .001 |
| 216 | Worrying about things said that hurt others feelings. | .63 | .06 | .001 |

Notes: RC7 = Dysfunctional Negative Emotions; Estimate = Standardized Estimate; S.E. = Standard Error; *p* = probability. All items contribute to the total score. If they are endorsed true, with the exception of those marked with an *, which contribute to the total score when endorsed false.

Table A8*Standardized Parameter Estimates: RC8*

| Items | Abbreviated Item Description | Estimate | S.E. | <i>p</i> |
|-------|--|----------|------|----------|
| 27 | Soul leaves body. | .63 | .10 | 0.001 |
| 49 | Attacks in which can't control body but knows what's going on. | .37 | .10 | 0.001 |
| 68 | Something was making me do things through hypnotizing me. | .56 | .11 | 0.001 |
| 108 | Bothered by hearing strange things. | .59 | .09 | 0.001 |
| 142 | Someone controlling thoughts. | .77 | .10 | 0.001 |
| 204 | Someone tries to influence my mind. | .62 | .09 | 0.001 |
| 222 | Hear voices without knowing their location. | .84 | .09 | 0.001 |

Notes: RC8 = Aberrant Experiences; Estimate = Standardized Estimate; S.E. = Standard Error; *p* = probability. All items contribute to the total score. If they are endorsed true, with the exception of those marked with an *, which contribute to the total score when endorsed false. The model for RC8 was recalculated without Item 168, as it was a Heywood case. Item 168 = seeing things that others could not.

Table A9*Standardized Parameter Estimates: RC9*

| Items | Abbreviated Item Description | Estimate | S.E. | <i>p</i> |
|-------|--|----------|------|----------|
| 58 | Stood in the way of others because of principle. | .39 | .08 | .001 |
| 61 | Sleep did not seem necessary for days. | .51 | .08 | .001 |
| 86 | Without reason, I feel happy. | .62 | .07 | .001 |
| 94 | When bored, like to stir some excitement. | .49 | .08 | .001 |
| 124 | Becomes very excited. | .56 | .08 | .001 |
| 177 | Fits of uncontrollable laughing and crying. | .48 | .08 | .001 |
| 182 | Never done anything dangerous for thrill* | -.18 | .09 | .043 |
| 236 | So excited, find it hard to sleep. | .68 | .07 | .001 |

Notes: RC9 = Hypomanic Activation; Estimate = Standardized Estimate; S.E. = Standard Error; *p* = probability. All items contribute to the total score. If they are endorsed true, with the exception of those marked with an *, which contribute to the total score when endorsed false.