p-ISSN: 2184-2205 e-ISSN: 2184-3414 ISBN: 978-989-35106-6-7 © 2024

MMPI-A TEMPORAL STABILITY STUDY IN TWO SAMPLES OF PORTUGUESE ADOLESCENTS, WITH AND WITHOUT CLINICAL COMPLAINTS

Maria João Santos[1], Maria João Afonso[1], & Rosa Novo[1,2]
[1]Faculty of Psychology, University of Lisbon (Portugal)
[2]CICPSI (Research Center for Psychological Science) (Portugal)

Abstract

The present research is a temporal stability study of the Portuguese version of the Minnesota Multiphasic Personality Inventory for Adolescents (MMPI-A; Silva et al., 2006). This personality and psychopathology inventory, the Portuguese adaptation of the original version (MMPI-A; Butcher et al., 1992), was approved by the University of Minnesota Press. This work analysed the temporal stability of the personality and psychopathology measures provided by the MMPI-A, namely the Validity, Clinical, Content and Supplementary scales stability, in two samples contrasted by the participants' clinical condition. The study followed a repeated measures design, with a 7-day average interval between administrations. The overall sample included 146 participants aged between 14 and 18 years (M = 17.09; SD = 1.33) organized in two samples: a) Non-clinical Complaints sample (N = 62), including 11 males and 51 females; b) Clinical Complaints sample (N = 84), including 11 males and 73 females. The MMPI-A was administered in small groups and individual sessions, after informed consent by the participants, or their legal representatives, when under 18. Descriptive statistics (M and SD), Pearson correlation coefficients (r_{tt}), and paired-samples t-tests were used on five Validity scales, ten basic Clinical scales, fifteen Content scales, and six Supplementary scales. The temporal stability indices (r_{tt}) revealed highly significant values ($p < .001$), for all scales, in both samples. In the Non-clinical sample, the r_{tt} coefficients for the Validity scales varied between .83 and .90, for the Clinical scales, between .75 and .95, for the Content scales, between .75 and .93, and for the Supplementary scales, between .77 and .89. As for the Clinical sample, r_{tt} indices for the Validity scales varied between .75 and .91, for the Clinical scales between .70 and .92, for the Content scales, between .69 and 91, and for the Supplementary scales, between .75 and .88. Despite the high temporal stability indices, statistically significant differences were found between administrations, in a few scales. As should be expected, the MMPI-A Portuguese version's measures revealed good to very good temporal stability, within a one-week interval between assessments, both in non-clinical and clinical samples of Portuguese adolescents.

Keywords: *MMPI-A, temporal stability, test-retest, adolescent clinical sample, adolescent non-clinical sample.*

1. Introduction

Within the domain of Clinical Assessment, the present paper presents psychometric evidence for the Portuguese version of a multidimensional personality and psychopathology instrument: the MMPI-A (Butcher et al., 1992). This is one of the most internationally used in the psychological assessment of adolescents, especially in clinical (Archer, 2005) and forensic (Archer, et al., 2006) contexts. The results herein presented constitute a part of a larger project, aiming the adaptation and psychometric study of the MMPI-A in Portugal, considering the growing need for theoretically sound and psychometrically updated assessment tools, specifically conceived and improved for adolescent populations.

In the domain of psychological assessment, questionnaires and self-report instruments are widely used and play an essential role in personality and psychopathology evaluation. The Minnesota Multiphasic Personality Inventory – Adolescent (MMPI-A; Butcher et al., 1992) is used as a reference since the nineties, in different contexts, aiming at problem identification, diagnosis, and treatment planning. The MMPI-A can be described as a self-report personality and psychopathology inventory, that can be administered individually or in small groups to teenagers aged 14 to 18. The MMPI-A was

specifically constructed for adolescents, departing from the methodology used to develop the original MMPI designed for adults. It includes 478 true/false items organized into a diversity of scales assessing multiple dimensions, both of normal personality and of psychological and personality disorders. The MMPI-A has several Validity scales to detect response attitude patterns, such as defensiveness or responding inconsistency. The primary Clinical scales are the same as those included in the original MMPI and MMPI-2, and thus include: Hypochondriasis, Depression, Hysteria, Psychopathic Deviate, Masculinity-Femininity, Paranoia, Psychasthenia, Schizophrenia, Hypomania, and Social Introversion. Additionally, the instrument comprises several Content and Supplementary scales (Archer & Krishnamurthy, 2002).

Since personality traits and psychopathology stand out in research as stable and consistent psychological manifestations over short periods of time (Harris et al., 2016), their assessment must be consistent (reliable), displaying independence from error due to the moment or occasion of administration. Therefore, correlations between two occasions (test-retest) should be positive and highly significant, confirming the results' temporal stability, and few differences are expected for the measures, between the test and the retest mean values, over relatively short periods of time. This approach to stability is particularly important for an inventory used to identify psychological disorders or psychopathology when clinical complaints are present. The comparison of temporal stability indicators between samples with and without clinical complaints may, then, contribute to support the use of the MMPI-A for the assessment of personality and psychopathology in the whole adolescent population.

2. Objectives

With the aim of carrying out a reliability study of the MMPI-A in the Portuguese population, this paper presents temporal stability evidence for the personality and psychopathology measures provided by the MMPI-A, namely, the Validity, Clinical, Content and Supplementary scales, in two samples contrasted by the participants' clinical condition.

3. Methods

3.1. Participants

The overall sample included 146 participants aged between 14 and 18 (M = 17.09; SD = 1.33) and was organized in two samples: a) Non-clinical Complaints sample (N = 62), including 11 males (17,7%) and 51 females (82,3%) (M = 17.11, SD = 1.29); b) Clinical Complaints sample (N = 84), including 11 males (13,1%) and 73 females (86,9%) (M = 17.07, SD = 1.36).

As inclusion criterion for the Clinical sample, to be attending (or have attended before) psychological and/or psychiatric consultation services were considered, as reported by the participants in the MMPI-A Biographical Data sheet.

3.2. Instrument

The Portuguese adaptation (Silva et al., 2006) of the Minnesota Multiphasic Personality Inventory-Adolescent (MMPI-A; Butcher et al., 1992) is the object of this reliability study. It provides an array of Validity and Clinical scales, alongside some Content and Supplementary scales. The temporal stability analysis addressed these four groups of scales, including a total of 36 measures.

The scales raw scores are converted into uniform T-scores (M = 50; SD = 10) for eight of the MMPI-A basic scales (excluding scales 5 and 0) and, in general, a T $\geq$ 65 is considered clinically significant for these scales. The same procedure is used in the Content scales. For the Validity and the Supplementary scales, linear T scores are used.

3.3. Procedure

After approval by an ethical committee, this study followed a repeated measures design, with a 7-day average interval between administrations. The MMPI-A was administered in small groups or individual sessions, following the informed consent by the participants, or their legal representatives, when under 18.

The administrations were carried out by clinical psychologists, properly trained and following clear guidelines for administration setup and instructions. To avoid intentional efforts to memorize items or answers, in the first administration no information was given about the instrument to be replied in the second administration.

4. Results

Descriptive statistics (M and SD), Pearson correlation coefficients (r_{tt}), and paired-samples t-tests were used on the five Validity scales, ten basic Clinical scales, fifteen Content scales and six Supplementary scales, in the two samples' data. The results are presented in Table 1.

Table 1. Descriptive Statistics, Test-retest Correlations and Paired-Samples Comparisons between the T-scores (M = 50; SD = 10), for the Non-clinical and Clinical Complaints samples.

	Non-clinical Complaints sample (*N* = 62)							**Clinical Complaints sample** (*N* = 84)						
	Test		Retest		*t*	r_{tt}	*SEM* [a]	Test		Retest		*t*	r_{tt}	*SEM* [a]
	M	*SD*	*M*	*SD*	*(df = 61)*			*M*	*SD*	*M*	*SD*	*(df = 83)*		
Validity Scales														
L	57.48	11.03	58.74	11.22	-1.54	.83***	4.07	52.60	8.04	52.60	9.32	0.00	.75***	5.02
F	46.02	6.16	45.23	6.74	2.09*	.90***	3.21	54.57	8.46	54.39	9.25	0 44	.91***	2.93
F1	45.73	6.92	44.68	7.74	2.05*	.86***	3.81	54.26	8.89	53.27	9.39	2.08*	.89***	3.35
F2	46.61	6.06	46.29	6.46	0.79	.87***	3.59	54.35	9.09	54.68	9.81	-0.62	.87***	3.62
K	52.37	9.18	53.94	10.39	-2.33*	.86***	3.73	47.96	8.00	46.23	7.94	3.37**	.82***	4.20
Clinical Scales														
1.Hs	47.82	7.72	44.87	8.03	5.54***	.86***	3.76	66.24	11.31	63.65	12.44	3.61***	.85***	3.85
2.D	57.05	8.92	54.92	8.54	2.84**	.77***	4.78	72.99	9.07	71.65	10.36	1.80	.76***	4.86
3.Hy	49.55	8.61	47.60	9.04	3.12**	.85***	3.94	65.71	12.08	63.29	13.39	3.30**	.87***	3.67
4.Pd	46.82	7.62	45.29	6.19	2.70**	.81***	4.36	56.90	9.17	55.82	9.96	1.64	.80***	4.44
5.Mf	47.73	9.12	47.34	9.08	0.59	.84***	4.00	47.23	11.27	46.57	10.62	0.96	.84***	4.03
6.Pa	48.15	8.03	46.23	8.61	2.63*	.76***	4.86	59.13	9.60	57.01	10.83	2.45*	.70***	5.44
7.Pt	48.42	7.44	47.73	8.99	1.25	.88***	3.54	66.63	10.51	66.36	10.82	0.46	.87***	3.59
8.Sc	44.71	8.40	43.21	8.96	4.04***	.95***	2.35	59.61	11.50	58.23	12.31	2.25*	.89***	3.32
9.Ma	44.05	7.00	43.98	6.83	0.10	.75***	4.99	49.12	10.24	49.88	10.50	-1.13	.82***	4.22
0.Si	53.60	10.15	53.56	10.91	0.06	.93***	2.68	62.83	9.70	62.70	9.56	0.31	.92***	2.83
Content Scales														
A-anx	49.92	8.03	49.29	9.28	0.80	.75***	4.98	69.01	9.65	67.55	10.67	1.67	.69***	5.57
A-obs	48.82	8.27	48.19	9.10	0.89	.80***	4.48	61.10	10.37	62.52	10.67	-1.93	.79***	4.55
A-dep	48.65	8.32	47.08	9.21	2.87**	.89***	3.39	62.07	11.35	62.60	12.30	-0.91	.90***	3.13
A-hea	47.95	7.40	45.03	7.59	5.61***	.85***	3.86	63.81	11.54	61.43	11.78	3.27**	.84***	4.05
A-aln	47.97	8.44	46.53	8.52	2.23*	.82***	4.22	57.61	10.42	58.23	11.54	-0.94	.85***	3.85
A-biz	44.76	8.09	42.05	7.47	5.67***	.89***	3.38	54.63	10.60	51.89	10.32	4.36***	.85***	3.89
A-ang	45.44	9.34	44.16	10.93	1.71	.84***	3.96	52.32	10.39	54.73	11.93	-2.91**	.78***	4.70
A-cyn	46.29	8.16	47.06	9.58	-1.00	.77***	4.76	49.83	8.40	52.06	10.73	-2.83**	.74***	5.09
A-con	41.94	7.12	41.81	7.49	0.31	.90***	3.16	45.67	8.34	46.90	9.57	-2.29*	.86***	3.80
A-lse	49.24	8.48	48.08	9.87	1.57	.81***	4.36	62.57	13.02	62.94	13.18	-0.46	.84***	4.01
A-las	46.95	8.70	46.11	8.75	1.10	.76***	4.88	51.92	10.14	52.65	10.53	-1.14	.84***	4.04
A-sod	53.10	13.01	52.94	13.45	0.26	.93***	2.65	60.26	11.80	60.46	12.46	-0.32	.89***	3.33
A-fam	45.15	8.61	44.06	8.61	2.36*	.91***	2.97	53.25	11.10	52.92	11.83	0.62	.91***	3.02
A-sch	44.56	6.80	44.69	7.38	-0.26	.85***	3.90	52.46	11.11	52.27	10.82	0.37	.91***	3.07
A-trt	47.13	9.45	46.02	9.32	1.68	.85***	3.94	59.73	14.87	60.23	15.64	-0.56	.86***	3.80
Supplementary Scales														
MAC-R	47.50	8.33	46.61	8.08	1.25	.77***	4.81	50.92	8.97	51.82	9.55	-1.25	.75***	5.05
ACK	43.53	7.00	42.79	6.91	1.34	.80***	4.44	50.33	9.35	50.24	9.44	0.19	.88***	3.46
PRO	45.97	7.10	44.71	7.33	2.48*	.85***	3.91	50.49	8.59	50.40	9.66	0.17	.88***	3.41
IMM	43.35	6.86	42.61	7.45	1.39	.83***	4.14	52.20	9.32	52.44	9.70	-0.48	.88***	3.41
R	55.76	10.28	55.16	9.94	0.77	.82***	4.27	55.69	10.11	54.45	9.52	1.95	.83***	4.18
A	49.56	8.14	48.16	9.04	2.73**	.89***	3.26	62.21	7.89	62.17	7.99	0.11	.87***	3.55

Note. **Validity Scales:** L = Lie; F = Infrequency; F1 = Infrequency 1; F2 = Infrequency 2; K = Defensiveness. **Clinical Scales:** 1.Hs = Hypochondriasis; 2.D = Depression; 3.Hy = Hysteria; 4.Pd = Psychopathic Deviate; 5.Mf = Masculinity / Femininity; 6.Pa = Paranoia; 7.Pt = Psychasthenia; 8.Sc = Schizophrenia; 9.Ma = Hypomania; 0.Si = Social Introversion. **Content Scales:** A-anx = Anxiety; A-obs = Obsessiveness; A-dep = Depression; A-hea = Health Concerns; A-aln = Alienation; A-biz = Bizarre Mentation; A-ang = Anger; A-cyn = Cynicism; A-con = Conduct Problems; A-lse = Low Self-Esteem; A-las = Low Aspirations; A-sod = Social Discomfort; A-fam = Family Problems; A-sch = School Problems; A-trt = Negative Treatment Indicators. **Supplementary Scales:** MAC-R = MacAndrew-Revised; ACK = Alcohol / Drug Problem Acknowledgement; PRO = Alcohol / Drug Problem Proneness; IMM = Immaturity; A = Anxiety; R = Repression.

[a] *SEM*: Standard Error of Measurement.

*p < .05. **p < .01. ***p < .001.

The temporal stability coefficients (r_{tt}) revealed highly significant values (p < .001) for all scales, in both samples. Also in both samples, all but one of the coefficients were above .68, more than 50% of the coefficients (about twenty scales) were above the .85 threshold (Aiken & Growth-Marnat, 2006), and

about 15% were even above the .90 criterion (Gregory, 2015). In the Non-clinical Complaints sample, the r_{tt} coefficients for the Validity scales varied between .83 and .90, for the Clinical scales, between .75 and .95, for the Content scales, between .75 and .93, and for the Supplementary scales, between .77 and .89. As for the Clinical Complaints sample, r_{tt} indices for the Validity scales varied between .75 and .91, for the Clinical scales between .70 and .92, for the Content scales, between .69 and .91, and for the Supplementary scales, between .75 and .88. As a result of the high test-retest correlations, the standard error of measurement indices is relatively low when compared to the standard deviation of the T scores distribution (*Mdn* = 4 for both the non-clinical and clinical samples).

Despite the high temporal stability indices, statistically significant differences between administrations were found, only in a few scales, less than a half of the measures. In both samples, the second administration means were slightly lower than the first in the F and F1 Validity Scales (Infrequency), but only significantly in the Non-clinical sample, for the F scale. The opposite was observed for the K scale (Defensiveness) in the Non-clinical Complaints sample, but not in the Clinical Complaints sample, where the mean value of this scale significantly decreases in the second administration.

In the basic Clinical scales, a tendency was detected for the means to decrease, from the first to the second administration. In scales 1.Hs (Hypochondriasis), 2.D (Depression), 3.Hy (Hysteria), 4.Pd (Psychopathic Deviate), 6.Pa (Paranoia), and 8.Sc (Schizophrenia), significant differences were found between the two administrations in the Non-clinical Complaints sample, and in scales 1.Hs (Hypochondriasis), 3.Hy (Hysteria), 6.Pa (Paranoia), and 8.Sc (Schizophrenia), in the Clinical Complaints sample. For the Content scales, significant decreases revealed in A-dep (Depression), A-hea (Health Concerns), A-aln (Alienation), A-biz (Bizarre Mentation), and A-fam (Family Problems), in the Non-clinical Complaints sample, while in the Clinical Complaints sample, there were significant decreases in A-hea (Health Concerns) and A-biz (Bizarre Mentation), like in the Non-clinical Complaints sample, while some increases were found, for A-ang (Anger), A-cyn (Cynicism), and A-con (Conduct Problems). Finally, in the Supplementary scales, few significant differences between administrations were found, only decreases for the Non-clinical Complaints sample, specifically in the PRO (Alcohol / Drug Problem Proneness) and A (Anxiety) scales.

5. Discussion

The test-retest coefficients confirmed high temporal stability for all the 36 scales studied, leading to conclude that they provide reliable results, according to the psychometric literature requirements (AERA, APA & NCME, 2014; Aiken & Growth-Marnat, 2006; Gregory, 2015) and considering the expected consistency for a week interval, in the type of psychological constructs involved. Furthermore, since the results were of a similar magnitude in both samples, it may be concluded that the Portuguese version of the MMPI-A can be considered a highly reliable psychological measure, in what concerns the temporal error effect, both for adolescents with and without clinical complaints.

The range of the test-retest coefficients in the basic Clinical Scales (from .75 to .95 in the sample without clinical complaints, and from .70 to .92 in the sample with clinical complaints) are similar but slightly higher to those found in the original MMPI-A, since the test-retest correlations for these scales ranged between .65 (6.Pa) and .84 (0.Si) (Butcher et al., 1992). In a study with the Spanish version of the MMPI-A (Zubeidat, et al., 2011), the test-retest coefficients in the Clinical Scales varied from .71 (3.Hy) to .92 (7.Pt), in the Content Scales, from .82 (A-obs) to .91 (A-sod), and in the Supplementary Scales from .78 (MAC-R) to .81 (R). These results are very similar to those found in the present study.

Only a few differences in the paired-samples comparisons were observed in both samples, in less than half of the scales (15 and 11 scales out of 36, respectively, in the non-clinical and the clinical samples), generally representing a decrease from the first to the second administration. In the Clinical Scales, very significant decreases were found for Hypochondriasis (1.Hs) and Schizophrenia (8.Sc), and also significant results for Depression (2.D), Hysteria (3.Hy), Psychopathic Deviate (4.Pd), and Paranoia (6.Pa) scales, in the non-clinical sample. In the same way, significant decreases were detected in the sample with clinical complaints, for similar scales: Hypochondriasis (1.Hs), Hysteria (3.Hy), Paranoia (6.Pa), and Schizophrenia (8.Sc) scales. Somehow, response attitudes may impact the Clinical Scales results, thus the interpretation of these findings, especially on those scales with more items requesting the recognition of symptoms, must consider the change in response attitude between administrations. Even though the Infrequency scale (F) presented a statistically significant decrease only in the non-clinical sample, in both samples a tendency seemed to emerge in means for the lowering of the reporting of problems and difficulties, in the second administration (alongside with an increase of the K "defensiveness" scale, although only for the non-clinical sample). Then, the differences in response attitude are more pronounced in the non-clinical sample, which is the one where more significant

decreases were observed in clinical measures. While the mean differences are generally positive, depicting a decrease in reported symptoms and difficulties, it also suggests that familiarity with the inventory content may lead participants to minimize, in a second assessment moment, the experience of psychological problems and symptoms. It is noteworthy, in this context, the higher temporal stability of the mean results observed in the Clinical Complaints sample (less than one third of the scales displaying statistically significant differences).

In the Content scales, some differences between the two administrations were also found in both samples, but this time significant increases revealed in the "acting out" scales Anger (A-ang) and Conduct Problems (A-con), and in Cynicism (A-cyn), only for the Clinical Complaints sample. This result may be associated with the very significant decrease in defensiveness (K scale) only observed in this sample. On the other hand, in both samples, significant decreases in the Content scales have parallels with equivalent basic Clinical scales: in the non-clinical sample, the decrease between administrations for Depression

(A-dep), Health Concerns (A-hea), Bizarre Mentation (A-biz) and Alienation (A-aln) may be considered comparable to the decreases in the scales Depression (2.D), Hypochondriasis (1.Hs) and Hysteria (3.Hy), and Schizophrenia (8.Sc). And in the clinical sample, the Bizarre Mentation (A-biz) and the Health Concerns (A-hea) scales had significant decreases similar to the ones observed in Schizophrenia (8.Sc), Hypochondriasis (1.Hs) and Hysteria (3.Hy) Clinical scales. Finally, in the Supplementary scales, only two differences were found in the sample without clinical complaints, in the Alcohol / Drug Problem Proneness (PRO) and the Anxiety (A) scales, while the clinical sample displayed temporal stability in the means of all these scales.

Even though a shortcoming may be pointed out to this study, regarding the limited generalizability for its conclusions due to the reduced sample sizes, the methodological option of studying temporal stability departing from two samples contrasted by clinical condition, instead of one larger sample, revealed useful, bringing some valuable insights about the clinical merits of this inventory. The Portuguese version of the MMPI-A presented high to very high indicators of temporal stability, as expected for an inventory devised to assess stable constructs over time, as personality and psychopathology dimensions. These results contribute to support the use of the MMPI-A for the assessment of personality and psychopathology in the Portuguese adolescent population. Yet, the higher stability of the results obtained in the Clinical Complaints sample further highlight the value of this instrument for a reliable identification of psychological disorders or assessment of psychopathology, when clinical complaints are present.

References

Aiken, L. R. & Groth-Marnat, G. (2006). *Psychological testing and assessment* (12th ed.). Boston: Pearson.

American Educational Research Association (AERA), American Psychological Association (APA) & National Council on Measurement in Education (NCME) (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Archer, R. P. (2005). *MMPI-A: Assessing Adolescent Psychopathology* (3rd ed.). Lawrence Erlbaum Associates Publishers.

Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of personality assessment, 87*(1), 84–94. https://doi.org/10.1207/s15327752jpa8701_07

Archer, R. P., & Krishnamurthy, R. (2002). *Essentials of MMPI-A assessment*. New York: Wiley.

Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R. P., Tellegen, A., Ben-Porath, Y. S., & Kaemmer, B. (1992). *Minnesota Multiphasic Personality Inventory – A (MMPI-A): Manual for administration, scoring, and interpretation*. Minneapolis: University of Minnesota Press.

Harris, M. A., Brett, C. E., Johnson, W., & Deary, I. J. (2016). Personality stability from age 14 to age 77 years. *Psychology and aging*, *31*(8), 862–874. https://doi.org/10.1037/pag0000133

Gregory, R.J. (2015). *Psychological Testing: History, principles, and applications* (7th ed.). Pearson.

Silva, D., Novo, R., Prazeres, N., & Pires, R. (2006). *Inventário Multifásico de Personalidade de Minnesota (Adolescentes): Versão experimental portuguesa do MMPI-A.* Lisboa: Centro de Investigação em Psicologia e Faculdade de Psicologia e de Ciências da Educação da Universidade de Lisboa.

Zubeidat, I., Sierra, J. C., Salinas, J. M. & Rojas-García, A. (2011). Reliability and Validity of the Spanish Version of the Minnesota Multiphasic Personality Inventory-Adolescent (MMPI-A). *Journal of Personality Assessment, 93* (1), 26-32.