Paidéia
2023, Vol. 33, e3304.doi:https://doi.org/10.1590/1982-4327e3304
ISSN 1982-4327 (online version)

Psychological Evaluation

Temporal Stability of the Zulliger Test in Brazilian Adults

Jucelaine Bier Di Domenico Grazziotin[1]
Alexandre Pereira Tognon[2]
Silvana Alba Scortegagna[1]

Abstract: The reliability of a test obtained over time is an indispensable measure to ensure the use of the instrument. This study aims to explore the temporal stability of the Zulliger test in Brazilian adults. A total of 20 participants, aged 18 to 59 years, answered a sociodemographic questionnaire and the Zulliger in Comprehensive System (ZCS). Data collection took place from 2009 to 2019 (Test) and in 2021 (Retest). ANOVA did not show differences between the values of 90 (99%) ZCS variables analyzed between test and retest; 68% of the variables showed agreement between 0.40 and 1.00, and 18%. Reflex responses; human movement, animal movement; weighted sum of color responses; sum of animal + inanimate movements; mixed determinants; art and clouds contents; Responses to card I; felt stimulation; cooperative movement; sum of critical special codes; degree of control and stress tolerance denoted high and excellent stability levels intraclass correlation coeficiente (ICC ≥ 0,70). The ZCS demonstrates temporal stability, reliability, and psychometric properties that support its safe use.

Keywords: test validity, longitudinal studies, projective technique, psychological assessment, test reliability

Estabilidade Temporal do Teste de Zulliger em Adultos Brasileiros

Resumo: A fidedignidade de um teste obtida ao longo do tempo é uma medida indispensável para assegurar o uso dos instrumentos. Este estudo teve por objetivo explorar a estabilidade temporal do teste de Zulliger em adultos brasileiros. Vinte participantes, entre 18 e 59 anos de idade, responderam formulário de dados sociodemográficos e o Zulliger no Sistema Compreensivo (ZSC). A coleta de dados ocorreu de 2009 a 2019 (teste) e em 2021 (reteste). A ANOVA não mostrou diferenças entre os valores de 90 (99%) variáveis do ZSC analisadas entre o teste e reteste; 68% das variáveis apresentaram concordância entre 0,40 e 1,00 e 18%. Respostas reflexo; movimento humano, movimento animal; soma de cor ponderada; soma de movimentos animal+ inanimado; determinantes mistos; conteúdos Arte e nuvem; Respostas para cartão I, estimulação sentida; movimento cooperativo; soma códigos especiais críticos; grau de controle e tolerância ao estresse denotaram níveis de estabilidade alto e excelente Coeficiente de Correlação Intraclasse (ICC ≥ 0,70). O ZSC demonstra estabilidade temporal, confiabilidade e propriedades psicométricas que respaldam o seu uso com segurança.

Palavras-chave: validade do teste, estudos longitudinais, técnicas projetivas, avaliação psicológica, precisão do teste

Estabilidad Temporal del Test de Zulliger en Adultos Brasileños

Resumen: La confiabilidad de una prueba realizada a lo largo del tempo es una medida necesaria para asegurar el uso del instrumento. El objetivo de este estudio fue explorar la estabilidad temporal de la prueba de Zulliger en adultos brasileños. Veinte participantes, con edades entre 18 y 59 años, respondieron un formulario de datos sociodemográficos y el Zulliger en el Sistema Comprensivo (ZSC). La recolección de datos tuvo lugar de 2009 a 2019 (test) y en 2021 (retest). ANOVA no mostró diferencias entre los valores de 90 (99%) variables ZSC analizadas entre test y retest; El 68% de las variables mostró acuerdo entre 0,40 y 1,00, y 18%. Respuestas reflejo; respuestas de movimiento humano y animal, suma ponderada de color, suma de movimiento animal + inanimado; determinantes mixtos; Arte y contenido en la nube; Respuestas a la Tarjeta I; estimulación sentida; movimiento cooperativo; suma códigos especiales críticos; grado de control y tolerancia al estrés denotaron niveles de estabilidad altos y excelentes Coeficiente de correlación intraclase (ICC ≥ 0,70). El ZSC demuestra estabilidad temporal, confiabilidad y propiedades psicométricas que respaldan su uso seguro.

Palabras clave: validación de test, estudios longitudinales, técnicas proyectivas, evaluación psicológica, precisión de test

Psychological evaluation is considered an procedure aimed to understand the psychological functioning and behavior of an individual, thus playing a primary role in Empirical Psychology. When responding to demands from different contexts and providing information for decision making, the procedures should resort to the available research and to the selection of instruments with evidence of their validity and reliability (Bornstein, 2017; Chnaider & Nakano, 2021; Villemor-Amaral & Primi, 2009; Wechsler, Hutz, & Primi, 2019).

In Brazil, there are few technical resources that have demonstrated their empirical efficiency for evaluation

[1]*Universidade de Passo Fundo, Passo Fundo-RS, Brazil*

[2]*Hospital São Vicente de Paulo, Passo Fundo-RS, Brazil*

Article derived from the master's thesis that is being extended toward the development of a doctoral dissertation of the first author under the supervision of the third, defense stipulated for May 2023, in the Graduate Program in Human Aging – PPGEH of the University of Passo Fundo, state of Rio Grande do Sul. The study received financial support from the Coordination for the Improvement of Higher Education Personnel (CAPES) - Financing Code 001.

Correspondence address: Silvana Alba Scortegagna. Universidade de Passo Fundo. Bairro São Jose BR 285, Passo Fundo-RS, Brazil. CEP. 99.052-900. E-mail: silvanalba@upf.br

in various contexts and that are included in the list of favorable tests of the Psychological Test Assessment System (*Sistema de Avaliação de Testes Psicológicos* – SATEPSI). The Brazilian guidelines for the scientific parameters of psychological tests (Conselho Federal de Medicina [CFP], 2018) establish that every 15 years the tests need to present new normative parameters and studies to present evidence of validity and reliability. These guidelines seek to ensure a safe psychological assessment of the evaluated population (Grazziotin & Scortegagna, 2021; Wechsler et al., 2019).

Regarding psychological instruments, accumulating evidence of validity that ensures its relevance within the proposed uses and the yielded interpretations is a manner of ensuring that psychology can exercise its ethical and social commitment (CFP, 2018). Another indispensable measure to certify the quality of the results obtained is its reliability. A test is considered reliable when it is capable of reproducing consistent results over time or from different observers, indicating aspects such as coherence, accuracy, stability, equivalence, and homogeneity (Cohen, Swerdlik, & Sturman, 2014; Souza, Alexandre, & Guirardello, 2017).

Generally, psychological assessment instruments, such as the projective methods of Rorschach and Zulliger, have some similarities. These tests provide, indirectly and with the stimuli of unstructured ink stains, information on the psychological functioning of an individual throughout their life. The Zulliger test, specifically, has proven useful for answering questions about the structure and dynamics of personality, cognitive aspects, and the internal resources that a person has to deal with problems (Villemor-Amaral & Primi, 2009). Three Zulliger coding and interpretation systems are available with studies on its validity and standardization for use in Brazil: the Klopfer System (Vaz & Alchieri, 2016), the Paris School System (Resende & Nascimento, 2019), and the Comprehensive System (Villemor-Amaral & Primi, 2009). The published scientific articles related to Zulliger in Comprehensive system (ZCS; Cardoso, Gomes, Pacheco, & Viana, 2018) and the results from a search in several national and international databases (Scielo, PePSIC, IndexPsi Articles, Lilacs, PsycNET, and PubMED) show the psychometric properties of ZCS, from 2009 to 2021. The results corroborate the ZCS validity in standard application (Cardoso et al., 2018; Grazziotin & Scortegagna, 2018, 2021), based on the Rorschach Performance Assessment System – R-PAS (Hosseininasab et al., 2019) with a R-optimized management perspective (Gonçalves & Villemor-Amaral, 2020; Gonçalves, Zuanazzi, & Villemor-Amaral, 2019; Seitl et al., 2018; Villemor-Amaral & Gomes, 2020), all indispensable to assess the ZCS validity. However there is an expressive gap in temporal stability studies. Only one reliability study with longitudinal design was found (Villemor-Amaral, Machado, & Noronha, 2009) which shows the need to advance in the development of studies of this nature. Longitudinal studies allow a deeper understanding of the relationship between cause and effect of the observed variables (Hulley, Cummings, Brower, Grady, & Newman, 2015).

Aiming to fill this gap and to verify the accuracy of the ZCS test, Villemor-Amaral et al. (2009) conducted a temporal stability study (test-retest). The study was conducted with 25 men, non-patient, theology students, living in inland São Paulo, aged from 20 to 47 years (M = 28.3; SD = 7.7). The mean interval between test and retest was 5 months. Regarding protocols, 25% were blindly recoded by a rater, obtaining satisfactory accuracy indexes, from 60% to 100% (M = 85.42%). In the results, of the 16 indicators selected from the ZCS to perform correlation analyses using Pearson's method ($p \leq 0.05$), 14 (88%) presented satisfactory accuracy indices of 0.40 to 0.99 (reasonable to excellent). The variables [H:(H)+(Hd)+ Hd; W and CF] showed a 0.40 to 0.59 correlation (Reasonable stability). Other variables [M, (H), Hd] achieved a 0.60 to 0.74 correlation (moderate to high stability). More than half (n = 9; 56%) of the listed variables (R, S, D, Dd, M, C, (H) and (Hd); and H) obtained ≥ 0.70 correlation (high and excellent stability). The FC and EB variables did not present significant correlation.

Similar results in the investigation of temporal stability, but using the Rorschach Comprehensive System (CS), were found by Sultan, Adronikof, Réveillère, and Lemmel (2006) in a sample with 75 French non-patient adults, who agreed to perform the retest after three months. Nine variables showed correlations equal to or above 0.70 [*R, Zf, F, M, S, (2), lambda, EA, 3r+(2)/R*]. Lower values of temporal stability were justified by: (a) Distribution of variable values; (b) low intercoder reliability; (c) Low occurrence of some variables; changes when retesting because the individuals already knew the Rorschach method; (d) Less defensiveness of the evaluated in the retest; e) French sample present greater complexity, i.e., greater emotional distress and less perceptual adequacy.

Later, Sultan and Meyer (2009) continued the study of Sultan et al. (2006); their objective was to explore the impact of the frequency of the number of responses (R) on the stability levels of Rorschach-CS scores (Exner, 2003). We analyzed 83 variables from the lower section of the structural summary. The frequency of note Z (*Zf*), which had shown a high correlation, presented reduced stability when the mean number of responses was higher. The highest mean and R variation showed negative impact on the temporal stability of the Rorschach-CS variables.

Some studies with ZCS (Villemor-Amaral et al., 2009) and Rorschach-CS (Sultan et al., 2006) categorize the intervals between test and retest as short (average of 3 weeks up to 2 months); intermediate (average of 3 months up to 1 year); and long (average above 1 year). Considering a long time interval (above 3 years), which allows for changes in related variables, this study aims to explore the temporal stability of the Zulliger test in Brazilian adults. Thus, the following hypotheses were included:

Hypothesis 1: There will be no significant differences ($p \leq 0.05$) regarding the values obtained for each of the Zulliger variables, applied at different times. This hypothesis considered the findings of Sultan et al. (2006) and Villemor-Amaral et al. (2009).

Hypothesis 2: There will be level ranging from reasonable to excellent (0.40–1.00 ICC) of stability, for most variables (prevalence ≥ 0.10) of ZCS. This hypothesis was based on: (a) personality assessment studies (Cicchetti, 1994; Sultan et al., 2006; Villemor-Amaral et al., 2009) that consider stability level of ICC ≥ 0.75 as excellent; stability levels of 0.60 ≤ ICC < 0.75 as good; stability levels of 0.40 ≤ ICC < 0.59 as reasonable; and stability levels of ICC < 0.40 as poor; and (b) studies with ink stain tests (Sultan et al., 2006; Villemor-Amaral et al., 2009) that consider ICC ≥ 0.70 as having high level of stability.

Method

Participants

For convenience, 20 individuals (around 25% of the database) living in different neighborhoods of a municipality located inland state of Rio Grande do Sul, aged from 18 to 59 years participated in this reproducibility study; of which 13 were males (65%), with 7 to 13 years of formal study, the sample comprised store operators, sales attendants or promoters, and merchandising and food sector supervisors, most of which have been working in the same company since 2009, when the first ZCS data collection occurred.

The application of the Zulliger test (test-retest) in 20 individuals resulted in 40 protocols; the first data collection (test) occurred from July 2009 to August 2019 and the second data collection (retest) from March to July 2021. The inclusion criteria for the participants were: (1) completed the sociodemographic data questionnaire; (2) responded the Zulliger Test in the Comprehensive System (ZCS); (3) performed periodic medical and psychological examinations, and were considered able to work in commercial functions according to the Occupational Health Medical Control Program (*Programa de Controle Médico de Saúde Ocupacional* – PCMSO). Exclusion criteria for this study were: (a) individuals who at the time of data collection were absent from work, whether via medical recommendation, sick or maternal leave; (b) individuals who did not have Brazilian nationality.

Instruments

Sociodemographic data collection questionnaire. The questionnaire is composed of 16 questions, aiming to verify the inclusion or exclusion criteria and obtain information regarding age, gender, marital status, occupation, income in minimum wages (referring to the period collected), and health conditions of the sample.

Zulliger test in the Comprehensive system – ZCS (Villemor-Amaral & Primi, 2009). This instrument consists of a set of three ink stain cards (Card I, Card II, Card III) that have characteristics of ambiguity and incompleteness that incite the person to give different responses to the stimulus, allowing the investigation into personality characteristics. For this study, 117 variables listed in the structural summary (location, popularity, pair, determinants, contents, special codes, reasons and proportions) were selected.

Procedures

Data collection After obtaining the project approval by the Research Ethics Committee, data collection began. The 20 selected participants (25% of the total expected sample for the complete study) were invited to perform the retest and all agreed. This sample size (totaling 40 protocols) would already allow the performance of different statistical analyses and with results close to what is intended with the total sample.

The criterion used for the selection of the participants of this study was as follows: Two weekly shifts were made available (morning and afternoon) for the collection and scheduling of the retests in the pre-established days. The invitation to participate in the retest was sent to those on the list who had already performed the first application of the tests and continued working in the company. The participants were then scheduled for the retest application according to the availability of their schedule, in a specially designated location, considering the confidentiality of the data and respecting all preventive sanitary measures related to the COVID-19 pandemic.

The rapport was performed after participants signed informed consent form, and then, for the retest, the participants responded to the same instruments of the first application (Test), in the following order: Sociodemographic questionnaire (to verify the sample characteristics), followed by the ZCS, in an estimated time of 1 hour. The first author of this study conducted the applications/test-retest and discussed the codifications of the ZCS protocols with the advisor and third author of this study.

Data analysis. All the results generated by the data collection instruments of this research were entered in an Excel database. The analyses were performed using SPSS 27.0 program for Windows. Categorical variables were expressed as absolute and relative frequency and numerical variables as mean and standard deviation.

Initially, for reliability of the ZCS data, 25% of the protocols (n = 05/ test; 05/ retest) were referred for recoding by an independent rater, to then perform the analysis of the intraclass correlation coefficient (ICC – mixed model of two factors, absolute agreement type, 95% confidence interval). The interpretation of these results followed Cicchetti's (1994) recommendations, in which ICC ≥ 0.75 are considered excellent values; 0.60 ≤ ICC < 0.75 are good; 0.40 ≤ ICC < 0.59 are reasonable; and ICC < 0.40 are poor results.

Of the 95 variable encodings (Location, Z-note, popularity, determinant pair, contents, special codes) analyzed for the ZCS test, ICC analyses ranged from 0.77 to 1.00 (79%; n = 75), which are considered excellent, and from 0.63 to 0.70 (5%; n = 5), which are considered good. Other variables did

not present response records, zero variance (16%; n = 15). The five variables that presented good correlation (CF, FT, VF, FY, and Hh) were reviewed in the protocols and the raters entered into an agreement on the codifications.

After the reliability results of the protocols, the test and ZCS retest were analyzed. First, the normality parameters of the variables were verified using the Kolmogorov-Smirnov test. Subsequently, the differences between the test and retest of all ZCS variables were assessed using analysis of variance (ANOVA) models, in which the time effect (test vs retest) was specified as intra-subject effect and the time effect between the two measures as covariate. The p-values ≤ 0.05 were considered statistically significant. Next, the magnitude of differences between the values obtained in some variables (*R;R to Card I; R to Card II; R to Card III; F%; X-%; XA%; Xu%; Blends*) in the test and retest were observed in Bland-Altman graphs since the data of these variables can affect the result of all other variables.

Subsequently, the agreement between the values obtained in the test and in the retest was evaluated using intraclass correlation coefficient (ICC), also called reliability coefficient, of the consistent type, specifying the effects of measurement and subject as random, and considering the coefficients corresponding to a single measure. The mean differences and the ICC were presented with the respective 95% confidence intervals. According to Cicchetti's (1994) recommendations and the literature on ink stain instruments (Sultan et al., 2006; Villemor-Amaral et al., 2009) at long intervals (Exner, 2003; Sultan et al., 2006), the high level of stability for the analysis of ZCS variables was defined in this study as any level above 0.70, and ICC stability level of ≥ 0.75 were considered excellent .

Ethical Considerations

The study was approved by the Research Ethics Committee of the Universidade de Passo Fundo (CAAE No. 40295220.1.0000.5342). It includes Resolutions 510/2016 of the National Health Council and 09/2018, of the Federal Council of Psychology.

Results

Regarding sociodemographic variables, 13 (65.0%) of the 20 participants in this study were male. The mean age was 30.5 years (SD = 8.6) and 38.2 years (SD = 10.7) for the first and second evaluation, respectively. The mean time between evaluations was 7.7 years (SD = 3.8). Years of Schooling in both the first and the second evaluation was 10.8 years (SD = 2.4). The reported income for the first test was 1.9 (SD = 1.1) minimum wages, and in the retest was 3.1 (SD = 0.57) minimum wages. The functions performed by the study participants in the first evaluation were store operators (n = 9); sales attendants or promoters (n = 6), merchandising and food sector supervisors (n = 5) and, in the second evaluation, the study participants performed the functions of store operators (n = 5); attendants or sales promoters (n = 7), and merchandising and food sector supervisors (n = 8).

ZCS variables are distributed in three tables. Of these, 16 variables were related only with descriptive purpose: five of them (C, Tf, Sx, AB, M-) presented rare prevalence (< 0.05) and Regarding ZCS test and retest, of the 117 variables initially listed, 11 are not shown in the tables (T; V; Y; Ge; none; Cn; CP; DV; Fabcom; Alog; Contam) since they presented zero prevalence. The results of 106, 11 (DQV;C´F;VF;(Hd); Hx; (A); Bl;Id;PSV;DR;INCOM) unusual prevalence (≥ 0.05 and < 0.10). The data, therefore, will be addressed and discussed according to the results of the 90 variables that presented common prevalence (≥ 0.10). There was no significant impact of time on the progression of the variables.

For the distribution of variables in the tables, the order of the encodings and consequent interpretation was considered (R; R to Card I; R to Card II; R to Card III; W; D; Dd; S; DQ+; Cod; DQV;DQV/+; FQ+; FQo; FQu ; FQ-; WDA%; XA%; X-%; XU%; X+%; F%; Fpure; M; FM; m; FC; CF; C; FC´; C'F; FT; TF; FV; VF; FD; FY; SumC´; WSumC; SumT; SumV; Sumy; Fr+rf; PuroH; (H); (Hd); Hd; H.X; SumH; H:(H)+(Hd)+Hd; The; (A); Ad; An; Xy; An+Xy; Art, art, art. Ay; Bl; Bt; Cg; Cl; Phi; Food; Hh; Ls; In; Isolate, isolate, SC; Sx; Id; Afr; Zf ; Pop; Pair; 3r+(2); Blends; 2AB+(Art+Ay); AB; AG; COP; GHR; PHR; GPHR; MOR; PSV; PER; INCOM; DR; Sum6; WSum6; M-; Ma; Mp; the; p; S-; EB; AND; Fm+m; SumC´+T+V+Y; Es; noteD; Adjes; AdjD). Table 1 presents the test and retest results and the analysis of variance (ANOVA) and agreement (ICC) of 43 ZCS variables that represent the number of responses, location, quality of development, formal quality, and determinants.

The data obtained by ANOVA (Table 1) show that the variables did not present significant differences between the test and retest results. Of the 38 variables analyzed (prevalence ≥ 0.10), 60% (n = 23) showed agreement from 0.40 to 1.00. Of these, 32% (n = 12) presented reasonable ICC agreement of 0.40 to 0.59 (R; D; S; FQo; WDA%; XA%; X+%; F%; m; FC; CF; SumT); 18% (n = 7) good and high ICC agreement level of 0.60 to 0.74 (Rlam 01; RLam 03, W, DQ+, DQo, FT, FD); and 10% (n = 4) showed excellent ICC agreement ≥ 0.75 (Fr+rf; M, FM, WSumC).

Table 2 below presents the test and retest results and the analysis of variance (ANOVA) and agreement (ICC) of 28 ZCS variables representing the Contents.

The analyses of variance show that the analyzed variables did not present significant differences between the test and retest results. In the ICC agreement analyses (column 5), of the 22 analyzed variables (frequency ≥ 0.10), 73% (n = 16) showed agreement ranging from 0.40 to 1.00. Of these, 45% (n = 10) presented reasonable ICC agreement from 0.40 to 0.59 (PureH; Hd; SumH; H: (H)+ (Hd)+Hd; The; An; Xy; An+Xy; LS; Isolate), and 27% (n = 6) showed good and high ICC agreement of 0.60 to 0.74 (Cg, Food, Cl, Hh, Sc, Art).

Table 3 shows the test and retest results and the analysis of variance (ANOVA) and agreement (ICC) of 35 ZCS variables that represent popularity, even, ratios, proportions, and special codes.

Table 1

Differences and agreements in Zulliger (answers, location, development quality, formal quality, and determinants)

ZCS variables	Test (n = 20)		Retest (n = 20)		Difference* (95%CI)	ANOVA	ICC (Difference 95%)	Agreement
	M	*SD*	*M*	*SD*		*p*		
R	9.25	1.50	10.05	2.35	−0.80 (−1.66–0.06)	0.067	0.56 (0.17–0.80)	Reasonable
RCardI	3.10	0.79	3.30	0.87	−0.20 (−0.50–0.10)	0.172	0.72 (0.42–0.88)	High
RCardII	2.95	0.69	3.40	0.99	−0.45 (−0.97–0.06)	0.082	0.03 (−0.41−0.46)	Poor
RCardIII	3.20	0.89	3.40	0.99	−0.20 (−0.60–0.20)	0.304	0.61 (0.24–0.83)	Good
W	2.00	1.20	2.20	1.26	−0.20 (−0.67–0.27)	0.384	0.66 (0.33–0.85)	Good
D	6.25	1.94	7.00	1.38	−0.75 (−1.62–0.12)	0.086	0.43 (−0.01−0.73)	Reasonable
Dd	0.85	0.88	0.80	1.15	0.05 (−0.54–0.64)	0.861	0.17 (−0.28−0.56)	Poor
S	1.35	0.93	1.05	0.89	0.30 (−0.12–0.72)	0.150	0.49 (0.07–0.76)	Reasonable
DQ+	2.05	1.60	2.50	1.47	−0.45 (−1.05–0.15)	0.134	0.61 (0.24–0.83)	Good
DQo	6.85	2.21	7.00	3.06	−0.15 (−1.23–0.93)	0.773	0.65 (0.30–0.85)	Good
DQV***	0.20	0.41	0.05	0.22	0.15 (−0.08–0.38)	0.190	0.10 (−0.51–0.35)	***
DQV/+	0.15	0.37	0.10	0.31	0.05 (−0.18–0.28)	0.656	0.14 (−0.54–0.31)	Poor
FQ+	0.15	0.37	0.20	0.41	−0.05 (−0.34–0.24)	0.723	0.21 (−0.59–0.25)	Poor
FQo	4.85	2.13	5.45	2.01	−0.60 (−1.67–0.47)	0.252	0.42 (−0.02–0.72)	Reasonable
FQu	2.90	1.55	3.00	1.89	−0.10 (−1.33–1.13)	0.866	0.16 (−0.55–0.29)	Poor
FQ-	1.35	1.09	1.45	1.15	−0.10 (−0.81–0.61)	0.771	0.27 (−0.18–0.63)	Poor
WDA%	75.53	16.69	79.50	22.77	−3.97 (−14.05–6.10)	0.418	0.40 (−0.05–0.70)	Reasonable
XA%	82.40	13.91	85.90	11.83	−3.50 (−10.00–3.02)	0.274	0.45 (0.02–0.74)	Reasonable
X-%	17.16	13.38	13.53	11.11	3.63 (−3.07–10.32)	0.091	0.35 (−0.10–0.68)	Poor
XU%	33.61	13.71	22.40	9.89	11.22 (22.22–6.54)	0.310	0.16 (−0.29–0.55)	Poor
X+%	50.14	14.18	57.03	19.43	−6.89 (−1.22–15.02)	0.270	0.44 (0.01–0.73)	Reasonable
F%	45.04	17.04	45.57	10.20	0.53 (−7.70–6.70)	0.217	0.41 (−0.03–0.72)	Reasonable
F pure	4.10	1.45	4.60	1.67	−0.50 (−1.32–0.32)	0.879	0.34 (−0.11–0.67)	Poor
M	0.75	0.91	1.05	0.94	−0.35 (−1.17–0.47)	0.379	0.75 (0.47–0.89)	Excellent
FM	1.05	1.10	1.20	1.24	−0.30 (−0.61–0.01)	0.057	0.84 (0.63–0.93)	Excellent
m	0.55	0.76	0.40	0.68	0.15 (−0.47–0.17)	0.343	0.47 (0.04–0.75)	Reasonable
FC	0.95	0.82	1.05	0.60	−0.10 (−0.51–0.31)	0.616	0.41 (−0.03–0.71)	Reasonable
CF	0.90	0.64	0.70	0.73	−0.10 (−0.48–0.28)	0.587	0.49 (0.07–0.76)	Reasonable
C**	0.00	0.00	0.10	0.30	0.20 (−0.12–0.52)	0.211	**	**
FC′	0.35	0.59	0.40	0.60	−0.10 (−0.25–0.05)	0.169	0.33 (−0.12–0.67)	Poor
C′F	0.05	0.22	0.10	0.31	−0.05 (−0.38–0.28)	0.755	0.66 (0.31–0.85)	Good
FT	0.30	0.57	0.20	0.52	−0.05 (−0.15–0.05)	0.330	0.67 (0.33–0.85)	Good
TF**	0.05	0.22	0.00	0.00	0.10 (−0.12–0.32)	0.343	**	**
VF***	0.05	0.22	0.05	0.22	0.05 (−0.24–0.34)	0.722	0.05 (−0.48–0.39)	***
FV	0.15	0.37	0.10	0.45	0.05 (−0.06−0.16)	0.338	0.10 (−0.35−0.51)	Poor
FD	0.85	0.88	0.85	0.99	0.00 (−0.16–0.16)	1.000	0.64 (0.28–0.84)	Good
FY	0.35	0.59	0.15	0.37	0.20 (−0.36–0.36)	0.830	0.21 (−0.25–0.59)	Poor
SumC′	0.40	0.50	0.55	0.60	0.20 (−0.10–0.50)	0.174	0.27 (−0.18–0.63)	Poor
WSumC	1.30	0.80	1.28	0.79	−0.15 (−0.47–0.17)	0.335	0.81 (−0.57–0.92)	Excellent
SumT	0.35	0.59	0.25	0.72	0.03 (−0.22–0.27)	0.830	0.52 (0.11–0.78)	Reasonable
SumV	0.20	0.41	0.15	0.49	0.10 (−0.21–0.41)	0.505	0.10 (−0.35–0.51)	Poor
SumY	0.40	0.60	0.15	0.37	0.05 (−0.25−0.34)	0.720	0.39 (−0.06–0.70)	Poor
Fr + rf	0.15	0.37	0.15	0.37	0.00 (−0.15–0.15)	1.000	1.00 (−0.15−0.35)	Excellent

Note. *Adjusted difference for time between test and retest; ** Rare prevalence (Zero variance); *** Unusual prevalence. Positive values show that the test result was higher and negative values mean that the retest result was higher; Significance $p \leq 0.05$.

Table 2
Differences and agreements in Zulliger (contents)

ZCS variables	Test		Retest		Difference* (95%CI)	ANOVA	ICC (Difference 95%)	Agreement
	M	*SD*	*M*	*SD*		*p*		
Pure H	0.65	0.67	0.90	0.79	0.15 (−0.25–0.55)	0.444	0.52 (0.11–0.78)	Reasonable
(H)	0.15	0.37	0.40	0.60	−0.25 (−0.58–0.08)	0.128	0.17 (−0.28–0.56)	Poor
Hd	0.35	0.67	0.10	0.31	−0.25 (−0.56–0.06)	0.105	0.44 (0.01–0.74)	Reasonable
RH	0.05	0.22	0.15	0.37	0.25 (−0.01–0.51)	0.055	0.09 (−0.50–0.36)	***
Hx ***	0.05	0.22	0.15	0.49	−0.10 (−0.31–0.11)	0.339	0.06 (−0.48–0.39)	***
SumH	1.20	1.00	1.55	1.23	−0.10 (−0.34–0.14)	0.397	0.53 (0.13–0.78)	Reasonable
H:((H)+(Hd)+Hd	0.10	1.12	0.25	1.02	−0.35 (−0.86–0.16)	0.165	0.44 (0.04–0.73)	Reasonable
A	4.50	1.43	5.10	2.22	−0.15 (−0.70–0.40)	0.569	0.59 (0.21–0.81)	Reasonable
(A)***	0.05	0.22	0.10	0.31	−0.60 (−1.42–0.22)	0.141	0.07 (−0.50–0.37)	Poor
Ad	0.75	0.91	0.25	0.44	−0.05 (−0.24–0.14)	0.586	0.33 (−0.12–0.67)	Poor
An	0.70	0.57	0.70	0.98	0.50 (−0.12–0.88)	0.121	0.59 (0.21-0.82)	Reasonable
Xy	0.10	0.31	0.10	0.31	0.00 (−0.33–0.33)	1.000	0.44 (0.01–0.74)	Reasonable
An+Xy	0.70	0.73	0.80	1.00	0.10 (0.16–0.16)	1.000	0.46 (0.04–0.74)	Reasonable
Art	0.65	0.67	0.65	0.59	−0.10 (−0.54–0.34)	0.637	0.74 (0.44–0.89)	High
Ay	0.15	0.37	0.20	0.41	0.05 (−0.22–0.22)	1.000	0.14 (−0.31–0.54)	Poor
Bl***	0.10	0.31	0.05	0.22	−0.05 (−0.30–0.20)	0.673	0.08 (−0.49–0.37)	Poor
Bt	1.55	1.15	1.35	1.35	0.05 (−0.14–0.24)	0.586	0.31 (−0.15–0.65)	Poor
Cg	0.05	0.24	0.10	0.31	0.20 (−0.49–0.89)	0.542	0.66 (0.31–0.85)	Good
Cl	0.15	0.49	0.20	0.52	−0.05 (−0.06–0.16)	0.343	0.70 (0.38–0.87)	High
Fi	0.15	0.37	0.10	0.31	−0.05 (−0.24–0.14)	0.587	0.32 (−0.13–0.66)	Poor
Food	0.20	0.40	0.10	0.31	0.05 (−0.14–0.24)	0.582	0.64 (0.29–0.84)	Good
Hh	0.15	0.37	0.15	0.37	0.10 (−0.05–0.25)	0.165	0.61 (0.24–0.82)	Good
Ls	0.25	0.45	0.25	0.37	0.05 (−0.16–0.16)	1.000	0.47 (0.04–0.75)	Reasonable
Na	0.20	0.62	0.15	0.37	0.00 (−0.22–0.22)	1.000	0.29 (−0.17–0.64)	Poor
Isolate	2.45	1.43	2.35	1.73	0.05 (−0.22–0.32)	0.704	0.42 (−0.02–0.72)	Reasonable
SC	0.20	0.41	0.20	0.41	0.10 (−0.73–0.93)	0.802	0.69 (0.36–0.86)	Reasonable
Sx **	0.05	0.22	0.00	0.00	0.00 (−0.15–0.15)	1.000	**	**
Id ***	0.05	0.22	0.05	0.22	0.05 (−0.06–0.16)	1.000	0.05 (−0.47–0.39)	***

Note. *Difference adjusted for time between test and retest; **Rare prevalence (Zero variance); ***Unusual prevalence; Positive values show that the test result was higher and negative values mean that the retest result was higher; Significance $p \leq 0.05$.

Table 3

Differences and agreements in Zulliger (popularity, pair, ratios, proportions, and special codes)

ZCS variables	Test		Retest		Difference* (95%CI)	*p*	ICC (Difference 95%)	Agreement
	M	*SD*	*M*	*SD*				
Afr	0.46	0.12	0.49	0.09	−0.03 (−0.11-0.24)	0.342	0.16 (−0.56–0.29)	Poor
Zf	4.15	1.30	5.10	1.80	−0.95 (−1.70–0.20)	0.016	0.46 (0.03–0.75)	Reasonable
Pop	1.55	0.60	1.85	0.75	−0.30 (−0.62–0.02)	0.061	0.53 (0.13–0.78)	Reasonable
Pair (2)	2.70	1.26	3.10	0.97	0.25 (−0.02–0.52)	0.063	0.11 (−0.34–0.52)	Poor
3r+(2)	3.10	1.37	3.50	0.83	−0.40 (−0.19–0.99)	0.174	0.12 (−0.33–0.53)	Poor
Blends	1.35	1.09	1.20	1.15	−0.40 (−0.19–0.99)	0.174	0.70 (0.38–0.870)	High
2AB+ (Art+Ay)	0.55	0.76	0.90	0.91	0.00 (−0.16–0.16)	1.000	0.38 (−0.06–0.70)	Poor
AB **	0.00	0.00	0.10	0.45	−0.35 (−0.79–0.09)	0.117	**	**
AG	0.35	0.59	0.60	0.68	−0.10 (−0.31–0.11)	0.326	0.50 (0.08–0.76)	Reasonable
COP	0.30	0.47	0.50	0.61	−0.25 (−0.54–0.04)	0.084	0.71 (0.41–0.88)	High
GHR	0.90	0.73	1.20	0.89	−0.20 (−0.40–0.01)	0.326	0.65 (0.30–0.84)	Good
PHR	0.75	0.83	0.65	0.75	−0.50 (−0.83–0.17)	0.117	0.33 (−0.12–0.67)	Poor
GPHR	0.15	1.23	0.45	1.15	−0.10 (−0.51–0.31)	0.610	0.44 (0.01–0.73)	Reasonable
MOR	0.35	0.58	0.10	0.45	−0.30 (−0.88–0.28)	0.293	0.64 (0.28–0.84)	Good
PSV ***	0.10	0.31	0.05	0.22	0.25 (−0.04–0.46)	0.024	0.66 (0.31–0.85)	Good
PER	0.45	0.69	0.45	0.76	0.05 (−0.05–0.15)	0.311	0.50 (0.08–0.77)	Reasonable
INCOM ***	0.10	0.31	0.05	0.22	0.00 (−0.35–0.35)	1.000	0.66 (0.31–0.85)	***
DR ***	0.10	0.31	0.05	0.22	0.05 (0.05–0.15)	0.311	0.66 (0.31–0.85)	***
Sum 6	0.15	0.49	0.15	0.37	0.05 (0.06–0.16)	0.333	0.72 (0.41–0.88)	High
WSum6	0.40	1.27	0.35	0.88	0.00 (−0.14–0.14)	1.000	0.72 (0.41−0.88)	High
M-**	0.00	0.00	0.05	0.22	0.05 (−0.32–0.42)	0.781	**	**
Ma: Mp	0.15	0.75	0.25	0.97	−0.20 (−0.48–0.08)	0.155	0.51 (0.10–0.77)	Reasonable
Ma	0.45	0.60	0.55	0.69	0.15 (−0.21–0.51)	0.390	0.76 (0.49–0.90)	Excellent
Mp	0.30	0.58	0.50	0.69	−0.10 (−0.32–0.12)	0.341	0.53 (0.12–0.78)	Reasonable
a	1.05	1.05	1.35	1.04	−0.30 (−0.80–0.20)	0.218	0.18 (−0.07–0.49)	Poor
p	1.40	1.50	1.35	1.23	0.05 (−0.43–0.53)	0.830	0.22 (−0.35–0.02)	Poor
S-	0.20	0.52	0.10	0.45	0.10 (−0.24–0.44)	0.548	0.09 (−0.50–0.36)	Poor
EB	−0.62	1.20	−0.48	1.33	−0.05 (−0.16–0.06)	0.339	0.62 (0.26–0.83)	Good
EA	2.17	1.06	2.60	1.52	−0.16 (−0.65–0.35)	0.538	0.63 (0.27–0.84)	Good
FM+m	1.55	1.43	1.55	1.10	−0.43 (−0.97–0.12)	0.118	0.77 (0.51–0.90)	Excellent
SumC′+T+V+Y	1.25	1.21	1.05	1.23	0.20 (−0.41–0.41)	1.000	0.66 (0.32–0.85)	Good
es	2.90	2.25	2.55	2.50	0.20 (−0.29–0.69)	0.396	0.70 (0.38–0.87)	High
Note D	−0.80	2.77	−0.17	2.50	0.35 (−0.38–1.08)	0.329	0.74 (0.44–0.89)	High
Adjes	1.75	1.99	1.93	1.62	−0.63 (−1.55–0.30)	0.175	0.59 (0.21–0.82)	Reasonable
AdjD	0.00	2.36	0.60	2.61	−0.18 (−0.97–0.62)	0.650	0.76 (0.49–0.90)	Excellent

Note. *Difference adjusted for time between test and retest; **Rare prevalence (Zero variance); *** Unusual prevalence; Positive values show that the test result was higher and negative values mean that the retest result was higher; Significance of $p \leq 0.05$.

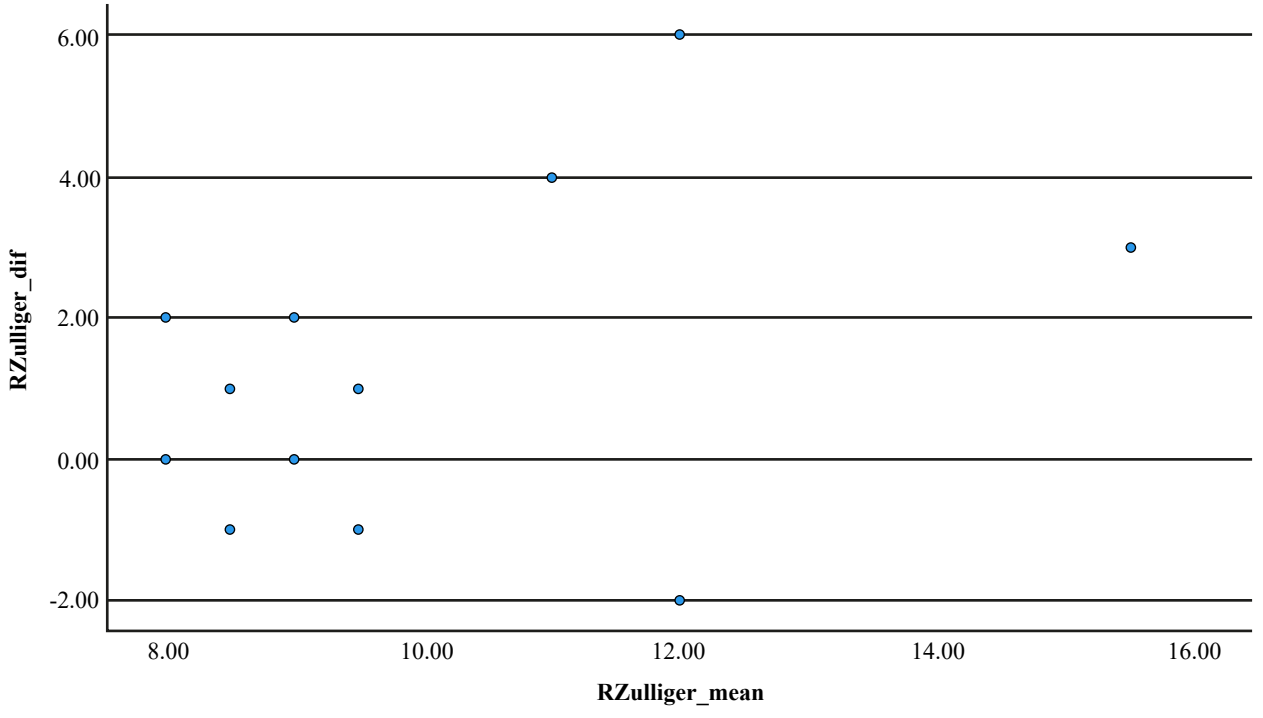
The data obtained by ANOVA show that the *ZF* variable showed a significant increase in the retest, the other variables did not show differences between results. Of the 30 analyzed variables (prevalence ≥ 0.10), 73% (*n* = 22) showed agreement of 0.40 and 1.00. Of these, 27% (*n* = 8) presented reasonable ICC agreement from 0.40 to 0.59 (Zf; Pop; Mp; Ma:Mp; AG; GPHR; PER; ADjes); 37% (*n* = 11) presented good and high ICC agreement of 0.60 to 0.74 (GHR, Mor, EB, EA, SumC´+T+V+Y, es; Blends; Cop; Sum6, WSum6; Note); and 10% (*n* = 3) exhibited excellent ICC agreement ≥ 0.75 (Ma, FM+m, AdjD).

Briefly, Tables 1, 2, and 3 show that the variables did not present differences between the test and retest results. There was agreement between 0.40 and 1.00 in 68% (*n* = 61) of the variables (reasonable to excellent). Of these, 33% (*n* = 30) with ICC ranging from 0.40 to 0.59; 27% (*n* = 24) with ICC from 0.60 to 0.74; and 8% (*n* = 7) with ICC ≥ 0.75 (Fr+rf; M, FM, Ma, WSumC, FM+m, AdjD). Composing 18% (*n* = 16) ofthe variables (Fr+rf; M; FM; Ma; WSumC; FM+m; AdjD; Rlam 01; Cl; Es; Blends; Art; Cop; Sum6; WSum6; NoteD) with ICC ≥ 0.70, of high and excellent level of stability.

The Bland-Altman plot reveals that, while the R variable may appear concordant (most values are close to zero and within the 95% confidence interval), it indicate some variation with an increased number of responses. The outer limit of the confidence interval suggests small-magnitude differences. Such variation is more visible on Card II . Regarding the other variables, we observed no correlation between the difference and the amplitude of the values. Therefore, Figure 1 shows the description of the number of responses.

Figure 1

Bland-Altman plot for the total number of responses



Discussion

The temporal stability of a test has special relevance for performing psychological evaluations. Firstly, from a psychometric perspective, large correlations for short-term test-retest are necessary to be confirm the reliability of the measures used (Sultan et al., 2006; Villemor-Amaral et al., 2009). Secondly, reasonable stability in long-terms should be expected in measures that are supposedly related to stable personality characteristics (Sultan et al., 2006); thus, this study aimed to explore the temporal stability of the Zulliger test in Brazilian adults.

In this study, we found no significant differences between the measurements in the test and retest for most of the variables listed in the ZCS (99%), of all the listed variables, only the ZF variable (*n* = 1;1%) showed a significant difference, denoting an increase in the retest. This difference may be due to other variables (R, W, DQ+, S). Thus, the Bland-Altman plot showed small oscillations in R, especially when individuals increased the number of responses and on Card II. Nevertheless, in its almost totality, the results confirm Hypothesis 1. Notably, the Zulliger test allows us to evaluate the structure or psychological functioning, the dynamics of personality, as well as the cognitive aspects and the internal resources to deal with problems (Villemor-Amaral & Primi, 2009). Thus, considering the long interval between the two applications (*M* = 7.7 years; *SD* = 3.8) individuals did not show significant changes in their way of thinking,

feeling, and solving problems. Individuals usually present variations in psychological dynamics during aging and due to internal and external factors (Chnaider & Nakano, 2021; Grazziotin & Scortegagna, 2021).

Nevertheless, in our study, the non-occurrence of differences between the results obtained in the ZCS at different times (test and retest) was expected, since the pairs had similar jobs, presented the same education level, and without occurrence of health issues. Validity studies that sought to compare independent groups, but with similar characteristics, reported some differences (Cardoso et al., 2018; Grazziotin & Scortegagna, 2021). Other studies, with the ZCS, that sought to compare independent groups with different characteristics or symptoms indicated several significant differences (Cardoso et al., 2018). Thus, the ZCS was relevant when answering questions from the psychological evaluation of Brazilian adults in similar or diverse situations. However, reliability studies (test and retest) with ZCS (Villemor-Amaral et al., 2009) are extremely rare. Considering the literature on psychological and personality evaluation (Cicchetti, 1994; Villemor-Amaral & Primi, 2009) and reliability studies with ink stain tests (Exner, 2003; Sultan et al., 2006; Sultan & Meyer, 2009), especially with the ZCS (Villemor-Amaral et al., 2009), we expected to find reasonable to excellent levels (ICC from 0.40 to 1.00) of temporal stability for most variables (prevalence ≥ 0.10) from ZCS. Thus, 68% (n = 61) of the variables listed showed agreement from 0.40 to 1.00 (reasonable to excellent). This data confirms Hypothesis 2, in which there would be temporal stability for most variables of the ZCS (prevalence ≥ 0.10) from ZCS.

The higher the indicators of agreement, the greater the possibility of the variable reflecting temporal stability, that is, stable characteristics or personality traits over time. The lower the indicators of agreement, the greater the possibility of the variable reflecting emotional states or the unstable characteristics of personality (Exner, 2003; Sultan et al., 2006; Villemor-Amaral et al., 2009). Many ZCS variables showed reasonable levels of stability (R; D; S; FQo; WDA%; XA%; X+%; F%; m; FC; CF; SumT; PureH; Hd; SumH; H: (H)+ (Hd)+Hd; A; An; Xy; An+Xy; LS; Isolate; Zf; P; Mp; Ma:Mp; AG; GPHR; PER; ADjes). Several variables presented good levels (R to card III, W, DQ+, DQo, FT, FD; Cg, Food, Hh, Sc; GHR; MOR, EB, EA, SumC´+T+V+Y) and high levels (R to Card I; Cl; Es; Blends; Art, art, art. COP; Sum6; WSum6; Stability note), suggesting aspects of state and traits. Some variables showed excellent stability levels (Fr+rf; M, FM, Ma, WSumC, FM+m, AdjD) suggesting traits characteristics. The other variables showed poor levels of agreement, suggesting temporal instability. In our study, unlike the findings of Villemor-Amaral et al. (2009) with the ZCS, but similar to the findings of Sultan et al. (2006) with Rorschach, few variables presented high level of stability, i.e., agreement above 0.70 (n = 16; 18%).

These results, as well as stability at reasonable levels (n = 30; 33%) for most variables, was expected for several reasons:

(1) The extensive number of analyzed variables (n = 90) and the number of responses generated in the protocols. Thus, the agreement of the number of responses (R↑T2; ICC = 0.56/reasonable) especially in R to card II (ICC 0.03/poor), seems to suggest that participants with lower motivation levels (R↓ and Fpure↑) may be more prone to change especially when approaching the task for the second time, become less defensive, and any changes in the amount of responses affect the other variables (Exner, 2003; Sultan et al., 2006; Sultan & Meyer, 2009). ZCS applications with an R-Optimized administration perspective (Gonçalves et al., 2019; Gonçalves & Villemor-Amaral, 2020; Seitl et al., 2018; Villemor-Amaral & Gomes, 2020) may minimize this problem. Furthermore, the results of many variables are considered according to their absolute values.

(2) Although the individual analyses of each variable provide relevant data, they also need to be considered interpretively, especially in longitudinal studies (Sultan et al., 2006; Villemor-Amaral et al., 2009). Ink stains tests, such as Rorschach and Zulliger, are complex instruments that consider the psychicological functioning and underlying characteristics of the individual (Gonçalves & Villemor-Amaral, 2020; Hosseininasab et al., 2019). Thus, the variables in Table 3 (ratios, proportions, and special codes) showed good and high agreements (ICC = 0.60 to 0.74/ Sum C´+T+V+Y; EA, EB, es, Note D, Sum6, Wsum6) and also excellent correlation level (ICC ≥ 0.75/ Fm + m + AdjD).

The variables SumV, SumY, SumC' (Table 1), when analyzed separately, had poor levels of agreement (ICC < 0.40) while SumT presented reasonable agreement (ICC = 0.52); when observing the SumV+Y+C´+T variables together, however, good agreement was obtained (ICC = 0.66). A person can express strong negative feelings for a higher Sum C' at baseline, the first test, and a higher Sum V or SumY in the retest. Moreover, ideas and needs can be expressed by m (ICC = 0.47) in the test and, mostly, by FM (ICC = 0.84) in the retest, i.e., FM+m has excellent agreement levels (ICC = 0.77). Accordingly, FC and CF presented reasonable agreement (ICC = 0.41 and 0.49); but when analyzed together (WSumC), the values reached were excellent (ICC = 0.81). This reasoning is corroborated by other authors (Exner, 2003; Sultan et al., 2006).

(3) Participants may express themselves differently between the two occasions, being more or less specific in verbalizing the contents and, consequently, causing an oscillation in the quality of the form (FQ+; FQo; FQu and FQ-) or content. The intermediate and long period of application can cause events and more changes related to internal (e.g., relationships; cognition) and external (e.g., age; wage; pandemics), which may affect emotional state and/or psychological functioning (Exner, 2003; Grazziotin & Scortegagna, 2021; Sultan et al., 2006).

Psychological evaluation plays an ethical and social function in answering questions from diverse contexts, situations, and populations, when based on instruments that ensure its psychometric properties with evidence of validity

and reliability (Bornstein, 2017; CFP, 2018; Chnaider & Nakano, 2021; Grazziotin & Scortegagna, 2018, 2021; Wechsler et al., 2019). In our study, the ZCS demonstrated reliability and indicated temporal stability. This is because reliability is conferred to a test when it is able to reproduce consistent results over time and indicate aspects such as coherence, accuracy, and stability (Cohen et al., 2014; Souza et al., 2017).

Future research with more extensive samples, shorter application times, and with R-Optimized and non-optimized application should be stimulated, allowing these findings to be confirmed or challenged. Our study contributes to a greater understanding of the relationship between cause and effect of the studied variables, the results found can serve as guidelines for other research on the psychometric properties of the Zulliger test and assist in the processes of psychological evaluation, while also collaborating to fill the literature gap and to foster scientific progress on the subject and stimulate the conduction of new research for the improvement of the scientific advancement.

References

Bornstein, R. F. (2017). Evidence-based psychological assessment. *Journal of Personality Assessment, 99*(4), 435-445. doi:10.1080/00223891.2016.1236343

Cardoso, L. M., Gomes, G. V. A., Pacheco, F. P., & Viana, J. L. D. (2018). Análise da produção de artigos científicos brasileiros sobre o Teste de Zulliger [Analysis of the production of Brazilian scientific articles on the Zulliger Test]. *Interação em Psicologia, 22*(3), 139-150. doi:10.5380/psi.v22i3.45821

Chnaider, J., & Nakano, T. C. (2021). Avaliação psicológica e envelhecimento humano: Revisão de pesquisas [Psychological assessment and human aging: Research review]. *Interação em Psicologia, 25*(3), 371-383. doi:10.5380/riep.v25i3.72089

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284-290. doi:10.1037/1040-3590.6.4.284

Cohen, R. J., Swerdlik, M. E., & Sturman, E. D. (2014). *Testagem e avaliação psicológica*: *Introdução a testes e medidas* [Testing and psychological assessment: Introduction to tests and measurements] (M. C. G. Monteiro, Trans., 8th ed.). Porto Alegre, RS: AMGH.

Conselho Federal de Psicologia. (2018, de 2 de maio). Resolução No. 9, de 25 de abril de 2018 [Resolution No. 9, 2018 April 25]. Estabelece diretrizes para a realização de Avaliação Psicológica no exercício profissional da psicóloga e do psicólogo, regulamenta o Sistema de Avaliação de Testes Psicológicos - SATEPSI e revoga as Resoluções No. 002/2003, No. 006/2004 e No. 005/2012 e Notas Técnicas No. 01/2017 e 02/2017. *Diário Oficial da União, seção 1.* Retrieved from https://www.in.gov.br/en/web/dou/-/resolucao-n-9-de-25-de-abril-de-2018-12526419

Exner, J. E., Jr. (2003). *The Rorschach: A comprehensive system*: *Vol. 1. Basic foundations and principles of interpretation*. New York, NY: Wiley & Sons.

Gonçalves, A. P., & Villemor-Amaral, A. E. (2020). Zulliger R-Optimized in the evaluation of depressive characteristics. *Paidéia (Ribeirão Preto), 30*, e3023. doi:10.1590/1982- 4327e3023

Gonçalves, A. P., Zuanazzi, A. C., & Villemor-Amaral, A. E. (2019). Aplicação R-Otimizada no Zulliger: Evidências de validade com pacientes depressivos [R-Optimized application in Zulliger: Evidence of validity with depressive patients]. *Avaliação Psicológica, 18*(2), 111-120. doi:10.15689/ap.2019.1802.16215.01

Grazziotin, J. B. D. D., & Scortegagna, S. A. (2018). Convergent validity of Zulliger-CS with the social skills inventory for the elderly. *Paidéia (Ribeirão Preto), 28*, e2832. doi:10.1590/1982-4327e2832

Grazziotin, J. B. D. D., & Scortegagna, S. A. (2021). Validity of Zulliger-SC in the cognitive assessment of elderly and long-lived adults. *Psico-USF, 26*(3), 571-583. doi:10.1590/141382712021260314

Hosseininasab, A., Meyer, G. J., Viglione, D. J., Mihura, J. L., Berant, E., Resende, A. C., … Mohammadi, R. M. (2019). The effect of CS administration or an R-Optimized alternative on R–PAS variables: A meta-analysis of findings from six studies. *Journal of Personality Assessment, 101*(2), 199-212. doi:10.1080/00223891.2017.1393430

Hulley, S. B., Cummings, S. R., Brower, W. S., Grady, D. G., & Newman, T. B. (2015). *Delineando a pesquisa clínica* [Outlining clinical research] (M. S. Duncan & A. G. Islabão, Trans., 4th ed.). Porto Alegre, RS: Artmed.

Resende, M. A., & Nascimento, E. (2019). *Z-SEP: Teste de Zulliger no Sistema Escola de Paris* [Zulliger Test in the Paris School System]. São Paulo, SP: Hogrefe.

Seitl, M., Dočkalová, J., Dostál, D., Kolařík, M., Hasoňová, V., Palová, K., & Fryštacká, K. (2018). Standardizace počtu odpovědí v Zulligerově testu: První krok při vývoji nového evidence-based systému [The Standardization of the number of responses in Zulliger test: The first step towards developing a new evidence-based system]. *E-Psychologie, 12*(4), 1-15. doi:10.29364/epsy.328

Souza, A. C., Alexandre, N. M. C., & Guirardello, E. B. (2017). Propriedades psicométricas na avaliação de instrumentos: Avaliação da confiabilidade e da validade [Psychometric properties in the evaluation of instruments: Evaluation of reliability and validity]. *Epidemiologia e Serviços de Saúde, 26*(3), 649-659. doi:10.5123/s1679-49742017000300022

Sultan, S., Adronikof, A., Réveillère, C., & Lemmel, G. (2006). A Rorschach stability study in a nonpatient adult sample. *Journal of Personality Assessment*, *87*(3), 330-348. doi:10.1207/s15327752jpa8703_13

Sultan, S., & Meyer, G. (2009). Does productivity impact the stability of Rorschach scores? *Journal of Personality Assessment, 91*(5), 480-493. doi:10.1080/00223890903088693

Vaz, C. E., & Alchiere, J. C. (2016). *Z-Teste Coletivo e Individual - Técnica de Zulliger* [Z-Test Collective and Individual - Zulliger technique]. São Paulo, SP: Hogrefe.

Villemor-Amaral, A. E., & Gomes, G. V. A. (2020). Comparação de indicadores patológicos e saudáveis do Zulliger com administração R-Otimizado [Comparison of pathological and healthy Zulliger indicators with R-Optimized administration]. *Psicologia: Teoria e Prática, 22*(3), 37-54. doi:10.5935/1980-6906/psicologia.v22n3p37-54

Villemor-Amaral, A. E., Machado, M. A. S., & Noronha, A. P. P. (2009). O Zulliger no sistema compreensivo: Um estudo de fidedignidade [The Zulliger in the Comprehensive System: A study of reliability]. *Psicologia: Ciência e Profissão, 29*(4), 656-671. doi:10.1590/S1414-98932009000400002

Villemor-Amaral, A. E., & Primi, R. (2009). *Z-SC: Teste de Zulliger no sistema compreensivo: Forma individual* [Z-SC: Zulliger test in the comprehensive system ZSC-Individually] São Paulo, SP: Casa do Psicólogo.

Wechsler, S. M., Hutz, C. S., & Primi, R. (2019). O desenvolvimento da avaliação psicológica no Brasil: Avanços históricos e desafios [The development of psychological assessment in Brazil: Historical advances and challenges]. *Avaliação Psicológica, 18*(2), 121-128. doi:10.15689/ap.2019.1802.15466.02

Jucelaine Bier Di Domenico Grazziotin is a PhD. candidate of the Graduate Program at the Human Aging at Universidade de Passo Fundo, Passo Fundo-RS, Brazil.

Alexandre Pereira Tognon is a PhD. in Epidemiology of the Hospital São Vicente de Paulo, Passo Fundo-RS, Brazil.

Silvana Alba Scortegagna is a Professor of the Graduate Program in Human Aging at the Universidade de Passo Fundo, Passo Fundo-RS, Brazil.

Authors' Contribution:
All authors made substantial contributions to the conception and design of this study, to data analysis and interpretation, and to the manuscript revision and approval of the final version. All authors assume public responsibility for the content of the manuscript.

Associate editor:
Adriana Martins Saur

Received: Mar. 30, 2022
1st Revision: Jul. 22, 2022
Approved: Oct. 19, 2022

How to cite this article:
Grazziotin, J. B. D. D., Tognon, A. P., & Scortegagna, S. A. (2023). Temporal stability of the Zulliger Test in Brazilian adults. *Paidéia (Ribeirão Preto), 33,* e3304. doi:https://doi.org/10.1590/1982-4327e3304