

INTRODUCTION

In recent years, computer-based testing (CBT) has grown in popularity over paper-and-pencil testing (P&P) and, in fact, several P&P tests are being “computerizing” (although they were standardized as P&P tests). Therefore it is of great importance to know the impact of this issue in the psychological assessment: Do psychometric properties differ? Are the scores interchangeable between P&P and CBT administration?

The focus of present research is to evaluate the score comparability of CB and P&P items in personality tests, specifically in four tests: BIP, (Hossiep & Paschen, 2006), TPT (Corral, Pamos, Pereña & Seisdedos, 2006), NEO-PI R (Costa & McCrae, 2008) and 16PF-5 (Russell & Karol, 2005).

HYPOTHESIS

- **Psychometric properties** of tests are the same in both administration methods: reliability and factor structure.
- **Scores** obtained from the tests are interchangeable regardless the administration method: DIF analysis, MANOVA.

RESULTS

RELIABILITY

Felt’s statistic (1969) was obtained to **compare the Cronbach’s alpha** coefficients in both conditions in the scales of the four tests. Significant values were obtained in several scales. The test with fewer significant values was TPT. However, these **differences were low**, and the means of these differences in the four tests were 0.0335, 0.0477, 0.0905 and 0.0272, respectively.

FACTOR VALIDITY

The configuration matrix were obtained from the factor analysis. The structures of TPT, NEO PI-R and 16 PF-5, similar to those contained in the respective manuals, were compared using the **congruence coefficient** (Wrigley and Neuhaus, 1955). Several authors suggest that there exists consistence if the values are higher than 0.80 or 0.90.

In all cases the coefficients were higher than 0.90, except NEO PI-R Extraversion factor, which value was 0.89. In most cases coefficients reached values **higher than 0.98**.

In the case of BIP, the correlation matrix were obtained for both samples, as shown below.

Correlación de Pearson		Motivation	Behaviour	Social skills	Psychic structure
eperfil	Motivation	1	,532**	,497**	,554**
	Behaviour	,532**	1	,722**	,776**
	Social skills	,497**	,722**	1	,764**
	Psychic structure	,554**	,776**	,764**	1
etea	Motivation	1	,499**	,531**	,612**
	Behaviour	,499**	1	,651**	,741**
	Social skills	,531**	,651**	1	,742**
	Psychic structure	,612**	,741**	,742**	1

** La correlación es significativa al nivel 0,01 (bilateral).

DISCUSSION

When comparing the structure of the tests in both administration methods it can be observed that **factor structure is not affected and that internal consistency differs these differences are not relevant**.

The **scores also differ** in both administration conditions, but **when the effect sizes were taken into account** rather than the critical values for signification, these **differences disappear or decrease**. The MANOVA results show that the lower effect size was found in TPT, and the higher one in NEO PI-R.

These results show that the differences between both administration methods are due mostly to the large sample size and not to real differences in the latent variables.

- REFERENCES
- Corral, S., Pamos, A., Pereña, J. & Seisdedos, N. (2006). *TPT. Test de personalidad de TEA. (3rd revised edition)*. Madrid: TEA Ediciones.
- Costa, P. T. & McCrae, R. R. (2002). *NEO PI-R. Inventario de Personalidad NEO Revisado. (3rd revised and extended edition)* (Spanish adaptation: A. Cordero, A. Pamos & N. Seisdedos). Madrid: TEA Ediciones.
- Feldt, L. S. (1969). A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two test. *Psychometrika* 34, 363-373.
- Hossiep, R. & Paschen, M. (2006). *BIP. Inventario Bochum de Personalidad y Competencias* (Spanish adaptation: D. Arribas, J. K. Woike & M. Wolff). Madrid: TEA Ediciones.
- Russell, M. T. & Karol, D. L. (2005). *16 PF-5. (8th revised and extended edition)* (Spanish adaptation: N. Seisdedos). Madrid: TEA Ediciones.
- Wrigley, C. % Neuhaus, J. O. (1955). *The matching of two sets of factors*. Contract Memorandum Report. University of Illinois.
- Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

SAMPLE

Sample consists of those examinees whose responses were stored in the TEA Ediciones scoring systems *e-perfil* and *e-teaediciones*. *e-perfil* allows the examiner to introduce the answers that the examinee gave in a P&P administration and get the scores and a profile. *e-teaediciones* is a computer based system where the examinee answers the same test.

	n		n
<i>e-perfil</i> (P&P)	17,754	BIP	2,063
	25,561	TPT	3,863
	3,810	NEO-PI R	747
	65,535	16PF-5	8,481
		<i>e-teaediciones</i> (CBT)	

DIF

To analyze the existence of differential item functioning, **logistic regression** was used as Zumbo (1999) proposes. Thus, effect size was obtained, as it was considered the most appropriate method taking into account the large sample sizes.

R² when predicting raw scores from the item responses was compared with R² when administration method was also introduced as independent variable (uniform DIF). Then, the first R² was compared with R² when the interaction between administration method and item responses was also introduced as independent variable (non-uniform DIF). **The effect size did not exceed 0.13** (value suggested by Zumbo) **at any item in the four tests**.

MANOVA

Multivariate Analysis of Variance was used to study the influence of the administration method in the scales of the tests. The result of the four **multivariate contrasts** (one for each test) was **statistically significant in all cases**. When the contrast for every single dependent variable (scale) was analyzed, it could be observed that exists significant differences in most of scales. However, **the effect size was very low**. In the table shown below it can be observed the number of scales with significant differences according to the analysis of variance and the average effect size (partial Eta square) for these scales in each test.

	BIP	TPT	NEO PI-R	16 PF-5
Number of variables with significant differences	11	13	27	14
Average effect size	0.0028	0.0012	0.0112	0.0025