

## MEASUREMENT INVARIANCE RESEARCH GROUP

Juana Gómez Benito  
Departament Metodologia de les Ciències del Comportament  
Facultat de Psicologia  
Universitat de Barcelona

One of the largest applied areas of statistical research within the behavioural sciences is undoubtedly that of psychometry, in which instruments are developed for measuring psychological traits. It is here that one finds the research group «Studies of measurement instrument invariance», whose perspective is that the appropriateness of psychological diagnoses depends directly upon the quality of the measures provided by such instruments and their fairness with respect to the different groups with which they are used.

**Psychometric techniques:** The measurement instrument must be a reliable and valid indicator of the trait being evaluated; the requirements which have to be met for this to be so are analysed at a theoretical level in Gómez (1989, 1996), particular emphasis being placed on the contribution which covariance structure models can make in terms of a more complex analysis of test reliability and validity. On the basis of this, new measurement instruments (Castro *et al.*, 1997; Puyuelo *et al.*, 1998) and computer programs (Martínez and Renom, 1993; Puyuelo *et al.*, 2002) are developed (Castro *et al.*, 1997), the quality of measurement provided by various existing scales is analysed (Forns and Gómez, 1990, 1994; Gómez and Forns, 1993*a*, 1993*b*; Stock, Okun and Gómez, 1994) and the equivalence of measures with respect to original versions is considered (Balluerka and Gómez, 2000; Balluerka *et al.*, 2000; Maydeu *et al.*, 2000).

**Analysing the effectiveness of techniques for detecting DIF:** The possible lack of test fairness with respect to certain variables (demographic, ethnic, cultural, etc.) would undermine validity. Differential item functioning (DIF) analysis aims to detect those items which may function differently for different groups, favouring some and being to the detriment of others. In order for such analysis to be accurate it is essential to use statistical techniques that offer higher rates of correct identifications and fewer classification errors. Simulation studies, in which various conditions, such as the amount of DIF, type of DIF, percentage of items with DIF, distribution of group ability, and sample size, are manipulated, can be used to compare the efficacy of the most widely used techniques for detecting DIF (Hidalgo and Gómez, 2000; Gómez and Navas, 2000) and the degree to which this efficacy may be optimized by applying iterative purification procedures to the trait measure (Navas and Gómez, 2001).

**Bias of items and/or tests:** Research in this area aims to elucidate if the differences found between groups reflect different levels of the trait being measured (impact)

or whether they are caused by systematic sources of variation unrelated to the trait (bias). In order to achieve this, the subgroups of items which are invariant with respect to possible bias variables must be identified across various instruments, as it is only these items which enable the measure to have the same meaning for all subjects. To this end, invariance studies are carried out at different levels: developmental (Gómez and Forns, 1996), linguistic (Ferrerres, González and Gómez, 2000), gender (Gómez and Navas, 1998) and questionnaire translation (Tomás, González and Gómez, 2000). In terms of techniques the following are applied: logistic regression, loglinear models, Mantel-Haenszel statistic, item response theory and confirmatory factor analysis. All these form part of what are known as conditional invariance methods, which match trait levels within groups so that they are comparable, thus enabling impact to be distinguished from bias.

## References

1. Balluerka, N. and Gómez, J. (2000). «Comparación entre los resultados obtenidos en la escala TDA-H (Trastorno por Déficit de Atención con Hiperactividad) en una muestra americana y en una muestra española de adultos». *Psicothema*, 12, 64-68.
2. Balluerka, N., Gómez, J., Stock, W. and Caterino, L. (2000). «Características psicométricas de las versiones americana y española de la escala TDA-H (Trastorno por Déficit de Atención con Hiperactividad): un estudio comparativo». *Psicothema*, 12, 629-634.
3. Castro, J., de Pablo, J., Gómez, J., Arrindell, W. A. and Toro, J. (1997). «Assessing rearing behaviour from the perspective of the parents: a new form of the EMBU». *Social Psychiatry and Psychiatric Epidemiology*, 32, 230-235.
4. Ferreres Traver, D., González Romá, V. and Gómez Benito, J. (2000). «Comparación del estadístico Mantel-Haenszel y la regresión logística en el funcionamiento diferencial de los ítems en dos pruebas de aptitud intelectual en un contexto bilingüe». *Psicothema*, 12, 214-219.
5. Forns Santacana, M. and Gómez Benito, J. (1990). «Factor Structure of the McCarthy Scales». *Psychology in the Schools*, 27, 111-115.
6. Forns-Santacana, M. and Gómez-Benito, J. (1994). «The cognitive, linguistic and adaptative development and academic achievement of pre-school children within the catalan immersion programme». In C. Laurén (ed.) *Evaluating Immersion Programs*. Vaasa: Tutkimuksia.
7. Gómez Benito, J. (1989). «Estimates of the internal consistency of a factorially complex composite». *Educational and Psychological Measurement*, 49, 571-578.
8. Gómez Benito, J. (1996). «Aportaciones de los modelos de estructuras de covariancia al análisis psicométrico». In J. Muñiz (ed.) *Psicometría*. Madrid: Universitat.

9. Gómez Benito, J. and Forns Santacana, M. (1993a). «Concurrent validity between the Columbia Mental Maturity Scales». *Perceptual and Motors Skills*, 76, 1177-1178.
10. Gómez Benito, J. and Forns Santacana, M. (1993b). «Confirmatory Factor Analysis of the McCarthy Scales». In R. Steyer, K. F. Wender and K. F. Widaman (eds.). *Psychometric Methodology*. Stuttgart: Gustav Fisher Verlag.
11. Gómez-Benito, J. and Forns-Santacana, M. (1996). «Factor structure of the McCarthy Scales in 7 years old Spanish Children». *Psychology in the Schools*, 33, 231-238.
12. Gómez, J. and Navas, M. J. (1998). «Impacto y funcionamiento diferencial de los ítems respecto al género en una prueba de aptitud numérica». *Psicothema*, 10, 717-728.
13. Gómez, J. and Navas, M. J. (2000). «A comparison of chi-square, RFA and IRT based procedures in the detection of DIF». *Quality & Quantity*, 34, 17-31.
14. Hidalgo Montesinos, M.<sup>a</sup>. D. and Gómez Benito, J. (2000). «Comparación de la eficacia de regresión logística polinómica y análisis discriminante logístico en la detección del DIF no uniforme». *Psicothema*, 12, 298-300.
15. Martínez, N. and Renom, J. (1993). *DEMOTAC: programa de ordenador para tests adaptativos computerizados*. Barcelona: PPU.
16. Maydeu-Olivares, A., Rodríguez-Fornells, A., Gómez-Benito, J. and D'Zurilla, T. J. (2000). «Psychometric properties of the Spanish adaptation of the Social Problem-Solving Inventory-Revised (SPSI-R)». *Personality and Individual Differences*, 29, 699-708.
17. Navas Ara, M. J. and Gómez Benito, J. (2001). «Effects of ability scale purification on identification of DIF». *European Journal of Psychological Assessment*.
18. Puyuelo, M., Renom, J., Solanas, A. and Wiig, E. (2002). *Evaluación del lenguaje: BLOC Screening. Manual de usuario*. Barcelona: Masson, S.A.
19. Puyuelo, M., Wiig, E., Renom, J. and Solanas, A. (1998). *Batería del lenguaje objetiva y criterial: Manual de usuario*. Barcelona: Masson, S.A.
20. Stock, W. A., Okun, M. A. and Gómez Benito, J. (1994). «Subjective well-being measures: reliability and validity among spanish elders». *International Journal of Aging and Human Development*, 38, 221-235.
21. Tomás Marco, I., González-Romá, V. and Gómez Benito, J. (2000). «Teoría de respuesta al ítem y análisis factorial confirmatorio: dos métodos para analizar la equivalencia psicométrica en la traducción de cuestionarios». *Psicothema*, 12, 540-544.