Personality and Individual Differences 82 (2015) 253–257

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

The relevance of measuring individual discrimination in personality: A study with the Test Anxiety Scale for Children

Pere J. Ferrando *

Research Centre for Behavioural Assessment (CRAMC), 'Rovira i Virgili' University, Spain

ARTICLE INFO

Article history:
Received 22 January 2015
Received in revised form 17 March 2015
Accepted 18 March 2015
Available online 30 March 2015

Keywords:
Personality measurement
Item Response Theory
Person discrimination
Person reliability
Test anxiety
Moderator effects

ABSTRACT

Personality theorists and researchers have long been aware that individuals respond to personality test items with different degrees of precision. However, standard models in common use in personality assume that the amount of individual discrimination is constant for all respondents. This article applies a simple Item Response Theory (IRT) model with an additional person discrimination parameter to a well-known anxiety measure: the Test Anxiety Scale for Children (TAS-C). The results show that the application is feasible and that the measurement of individual discrimination leads to important advantages both in terms of individual assessment (improved interpretation, differential precision of the estimates) and validity (differential predictability with respect to an external criterion of academic performance).

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

A historical literature review shows that personality measurement has not generally developed its own measurement models, but has adopted psychometric theories and models that were initially developed for measuring abilities. The most important of these are classical test theory (CTT), factor analysis (FA) and Item Response Theory (IRT) (see e.g., Ferrando, 1994). In general, this strategy has worked well and has produced many useful instruments (e.g., Eysenck & Eysenck, 1969; Reise & Waller, 2009). At the same time, however, in comparison to ability scores, personality scores generally have more measurement error, are more prone to response biases, and have weaker validity relations with relevant variables (Fiske, 1968; Ferrando, 2004). To sum up, even when the basis models used in personality allow reasonably good measures to be obtained, there is still ample room for improvement.

In all of the three frameworks mentioned above – CTT, FA, and IRT – the most general schema adopted in personality measurement is asymmetric. In the unidimensional case considered here, items are characterized by two relevant parameters: a difficulty or location parameter (estimated e.g., by the proportion of endorsement) which reflects the location of the item on the dimension or trait that is measured, and a discrimination parameter (estimated e.g., by the item factor loading) which reflects the quality of the item as a measure of this trait (Lord & Novick, 1968). Respondents, however, are only characterized by a single relevant parameter: the trait level of the individual (estimated e.g., by the test score or the factor score). Conceptually, this schema implies that all individuals are assumed to respond to personality items with the same degree of accuracy, sensitivity or discrimination (Ferrando, 2004; the three terms will be used interchangeably here). However, many personality theorists and researchers (Ferrando, 2004; Fiske, 1968; Guilford, 1959; Lanning, 1991; Taylor, 1977; Tellegen, 1988; Voyce & Jackson, 1977) have clearly stated that this is not the case. Some individuals respond to the set of items in a highly consistent, almost deterministic way, whereas others respond much more randomly. This result implies that a 'symmetric' modeling in which both items and respondents are characterized by different amounts of discrimination is more plausible for personality measurement than the modeling in common use.

In principle, the inclusion of an additional individual discrimination parameter in a psychometric model intended for personality items would have three main advantages. First, it would provide more information about the respondent. In this respect, it has been hypothesized that the accuracy with which the individual responds depends mainly on the relevance and degree of clarity and strength with which the trait is internally organized in him or her (e.g., Markus, 1977; Tellegen, 1988).

* Address: Universidad 'Rovira i Virgili', Facultad de Psicologia, Carretera Valls s/n, 43007 Tarragona, Spain. Tel.: +34 977558079.
E-mail address: perejoan.ferrando@urv.cat

http://dx.doi.org/10.1016/j.paid.2015.03.037
0191-8869/© 2015 Elsevier Ltd. All rights reserved.

Second, other things being constant, the standard error of the trait estimate would be necessarily smaller for reliable, highly discriminating individuals than for those who respond more randomly, and this result would have important consequences in clinical assessment and in selection or classification processes (e.g., Ferrando, 2014). Third, for both psychometric and conceptual reasons, individual discrimination would be expected to have a role of moderator in validity assessment. At the psychometric level, according to basic attenuation theory, the unreliability of test scores attenuates the validity coefficient (e.g., Lord & Novick, 1968), and the scores of the less discriminating individuals are less reliable. At the conceptual level, those individuals for whom the trait is relevant are expected to be more likely to display a stronger correspondence between trait self-description and relevant correlates (Markus, 1977; Paunonen, 1988). In both cases what is predicted then is that the most discriminating individuals would also tend to be the most predictable. Previous related research, however, suggests that, in practice, the differential validity effects would be modest at best (Lanning, 1991; Paunonen, 1988).

To date, personality researchers have proposed several CTT-based approximate descriptive procedures for assessing individual discrimination (Fiske, 1968; Lanning, 1991; Voyce & Jackson, 1977). More rigorous IRT-based proposals also exist in which both items and individuals can have different discrimination but their practical interest is only marginal. They were proposed either at the theoretical level or they are too complex to provide accurate and stable results with typical personality data sets (Ferrando, 2004; Lumsden, 1980). Recently, however, Ferrando (2014) has proposed a relatively simple and workable IRT model of this type intended for binary items that is thought to be appropriate for many personality applications.

1.1. Objectives

This article provides a conceptual, non-technical discussion of the IRT model mentioned above, and describes an application in which the model is used to analyze the responses to a well known personality measure. The purpose is twofold: illustrative and substantive. At the illustrative level the study aims to show how the model is applied, and, above all, how the results it provides are interpreted. At the substantive level the study aims to assess the extent to which the three main potential advantages of assessing individual discrimination – gains in information, differential accuracy in individual assessment, and differential gains in validity – are realized in practice.

1.2. Description of the model

The discussion of the model provided here is only conceptual and non-technical. Technically-oriented presentations can be found in Ferrando (2004, 2014).

The IRT model considered in this article is called D2PMM (dual two-parameter multiplicative model), and its basic equation is

$$P(X_{ij} = 1|\theta_i, \gamma_i, b_j, a_j) = \Phi(\gamma_i a_j(\theta_i - b_j)). \quad (1)$$

where Φ is the cumulative normal function. The D2PMM is a normal-ogive IRT model in which the probability that respondent i would endorse item j depends on two item parameters and two person parameters. The item parameters a_j (considered to be always positive) and b_j are, respectively, the item discrimination and the item difficulty. The person parameters γ_i and θ_i are the individual discrimination and the individual trait level. The difference $\theta_i - b_j$ is the signed person-item distance and is the primary response determinant. So, when the trait level of the individual dominates the item difficulty (i.e., when $\theta_i > b_j$) the probability of endorsing the item is greater than 0.5. The term "item difficulty" has been used so far because it is standard in IRT models, and reflects the fact, mentioned above, that these models were initially intended for measuring abilities. In the context of a self-reported personality scale, however, it might be better to refer to b_j as "item extremity" and this is the term that will be used below.

The person parameter γ_i (assumed to be positive) reflects the discriminating power of the individual, and moderates the sensitivity of the responding process based on the person-item distance via the product $\gamma_i a_j$. So, when both γ_i and a_j are high, the responding process becomes more deterministic. At the other extreme, when γ_i, a_j, or both approach zero, the process becomes more and more random, and the probability of endorsing the item approaches 0.5 no matter what the person-item distance is.

If the item discrimination is considered to be constant, the person discrimination parameter measures the extent to which the responses of the individual are sensitive to the location of the items on the trait continuum. In this way, an individual with a high value of γ_i is very reliable and mostly agrees with those items located below his/her trait level, and rejects those located above, thus producing a highly scalable pattern (Lanning, 1991). At the other extreme, a low value of γ_i means that the individual is highly unreliable and largely insensitive to item ordering. The resulting pattern of this individual is, therefore, almost random. Person discrimination is viewed as an individual-differences continuum, so all individuals are unreliable to some degree.

The D2PMM is identified by assuming trait θ to be distributed as a standardized variable (i.e., zero mean and unit variance) and the person discrimination γ to be distributed with unit mean. With these constraints, the items can be calibrated by fitting the standard two-parameter IRT model with any available IRT program. Next, in the scoring stage, the item discriminations and difficulties are taken as fixed and used for scoring individuals. For each respondent, the scoring results are the point estimates of θ_i (trait level) and γ_i (individual discrimination) together with the corresponding standard errors. For the moment, the scoring process must be carried out by using an ad-hoc program that can be obtained at no cost by requesting it from the author.

1.3. Current empirical study

In this article, the functioning of the D2PMM is assessed by re-analyzing a data set from a study by Ferrando, Varea, and Lorenzo (1999) based on a Spanish adaptation of the Test Anxiety Scale for Children (TAS-C). Given that the original study contains most of the relevant information, here only a summary of the results will be provided.

2. Method

2.1. Participants

The calibration sample was made up of 1022 primary school children (523 boys and 499 girls, between 12 and 14 years old) who were administered the TAS-C in classroom groups under standard instructions. Of this calibration sample, the validity study uses a sub-sample of 875 children for whom criterion scores were available.

2.2. Measures

The TAS-C version used in the study was a 30-item Spanish adaptation that used the same number of items, numerical index, and binary response format as the original American test (Sarason, Davidson, Lighthall, Waite, & Ruebush, 1960).

Download English Version:

https://daneshyari.com/en/article/890111

Download Persian Version:

https://daneshyari.com/article/890111

Daneshyari.com