Personality and Individual Differences 128 (2018) 1–6

Contents lists available at ScienceDirect

Personality and Individual Differences

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

Understanding item parameters in personality scales: An explanatory item response modeling approach☆

Hanyi Min*, Michael Zickar, Georgi Yankov

Department of Psychology, Bowling Green State University, United States

ARTICLE INFO

Keywords:
Item response theory
Item parameter
Personality measurement

ABSTRACT

Although item parameters are essential in psychometric frameworks, such as item response theory (IRT), few theories are available to guide their psychological interpretation. Moreover, the meaning of item parameters is generally harder to interpret for personality scales than for ability scales. The current study provides a comprehensive way to interpret personality item parameters, Generalized Linear and Nonlinear Models (GLNMMs). The GLNMMs model the effect of item features on psychometric properties, such as item discrimination and item location. Distinct from previous studies which use a two-step approach, the GLNMMs produce smaller standard errors of item features' coefficients, and allow examination of person covariates. The current study examines four item features – negative wording, subtlety, social desirability, and miscomprehension. In general, item discrimination negatively correlated with item subtlety, social desirability, and miscomprehension; item easiness positively correlated with subtlety and miscomprehension, and negatively correlated with negative wording.

1. Introduction

Well-designed personality studies require reliable and valid personality scales, and the building blocks of a personality scale are individual personality items. Although there are many guidelines on how to write good educational assessment items (e.g., Haladyna, 2012), systematic theoretical guidelines of personality item writing are lacking. Many researchers still perceive writing personality items as more of an art than a science (e.g., Kline, 2000; Ramsay & Reynolds, 2000). Item parameters in item response theory (IRT) can be a potential tool to develop guidelines in writing personality items, and a better understanding of item parameters in personality scales has implications for writing personality items. Unfortunately, researchers have limited understanding of parameters, especially in personality scales.

The goal of this study is to use a relatively new psychometric framework, generalized linear and nonlinear mixed models (GLNMMs; De Boeck & Wilson, 2005), to help researchers develop better insights into item parameter estimates. First, we briefly review previous studies focusing on the interpretation of item parameters. We compare parameter interpretations in personality scales with those in ability scales. Then we summarize work on GLNMMs. Applying the latter, we analyze how common item features influence item parameters. As will be explained, the advantage of GLNMMs is that item feature estimation and item parameter estimation are integrated in one step, thus facilitating interpretation. We hope that our study of understanding item features in personality items will help the shift away from the "art" of test development, where items were created based solely on content expertise, toward a more scientific approach of test development, where item generation is facilitated by the understanding of how item features associate with item psychometric properties.

1.1. Item parameter interpretation in personality tests

Unlike personality scales, item parameters in ability scales are reasonably intuitive to interpret. For example, in ability scales modeled with the commonly used three-parameter logistic (3PL) model, the a parameter indicates how discriminating an item is. The b parameter corresponds to difficulty, which relates to how cognitively challenging an item is. The third parameter, c, is related to guessing. It indicates the nonzero probability of observing correct answers for respondents with very low ability.

In personality measurement, however, the meaning of item parameters is much less understood and certainly less studied. Some of the concepts that have been used to describe the functioning of ability items do not translate well to personality items (see Zickar, 2001). A personality item cannot be described as easy or difficult, nor does it make sense to say, for example, that a respondent guessed the correct option on an extraversion item. As a result, it is relatively harder to interpret

☆ The IPIP respondents were archival data from the Eugene-Springfield community sample originally collected by Oregon Research Institute.

* Corresponding author at: Department of Psychology, Bowling Green State University, Bowling Green, OH 43403, United States.
E-mail address: mahnyi@bgsu.edu (H. Min).

https://doi.org/10.1016/j.paid.2018.02.012
Received 7 December 2017; Received in revised form 5 February 2018; Accepted 6 February 2018

item parameters for personality scales compared to ability scales.

A couple of studies have examined the interpretation of psychometric parameters for personality items. Rouse, Finger, and Butcher (1999) found that the average correlation between the *c* parameter estimate and the social desirability rating of that item across the *PSY-5* scales was 0.*34*. In contrast, Reise and Waller (2003) studied the c parameters using pathology and nonpathology direction MMPI items, and found that social desirability was unlikely to cause the non-zero *c* parameters; instead item-level multidimensionality might be the influencing factor. Davies, Norris, Turner, and Wadlington (2005) found that applicants had higher *c* parameters in transparent items compared to incumbents using the Hogan Personality Inventory (HPI; Hogan & Hogan, 1995). Zickar and Ury (2002) studied the interpretation of item parameter estimates for Goldberg's Adjective Checklist (Goldberg, 1992). They found that subtlety correlated negatively with item discrimination, *a*. Also, subtlety interacted with social desirability to affect item location, *b*.

Although these studies provided interesting results based on traditional IRT models, advances in IRT techniques provide a better psychometric framework for understanding the relationships between item features and personality item parameters.

1.2. Generalized linear and nonlinear mixed models (GLNMMs)

The previously mentioned efforts to interpret personality parameters adopted the same general two-step approach, in which item parameters are estimated in the first step, and then correlations are estimated between the parameter estimates derived in the first step and item features (e.g., social desirability and subtlety). Recent developments of methodology and analysis tools have led to a statistical modeling framework that combines these two steps into a single analysis: *explanatory item response modeling.*

The explanatory item response modeling is a more statistically appropriate modeling approach. The reason is that the standard errors for the estimated coefficients of item features are generally smaller than those estimated in the two-step approach (Embretson, 2010). When using the traditional two-step approach to estimate coefficients of item features, the sample size is the number of items in the focal scale. In contrast, when using explanatory IRT to estimate coefficients of item features, the sample size is responses from all participants. Thus, the estimated coefficients of items features have smaller standard errors and are more statistically justifiable. Another limitation of the two-step approach is that the two-step procedure assumes that there are no individual differences when interpreting item features. That is, person covariates cannot be estimated in the two-step approach, whereas person covariates can be modeled in the explanatory item response modeling to explain sources of individual differences.

The explanatory modeling framework is also called generalized linear and nonlinear mixed modeling (GLNMMs; De Boeck & Wilson, 2005). Many widely used models belong to the GLNMMs family, such as the 1PL, 2PL, and 3PL models. The linear predictor of GLNMMs can be expressed as

$$logit(P(y_{pi} = 1 \mid \theta_p)) = \sum_{k=0}^{K} \beta_k X_{ik} + \sum_{j=0}^{J} \alpha_j Z_{ij} \theta_p$$

where y_{pi} indicates person p's answer to item *i*. X_{ik} (k = 0, …, K) is the item specific covariate that can potentially influence item location (i.e., easiness) parameters; Z_{ij} (j = 0, …, J) is the item specific covariate that can potentially influence item discrimination parameters. β_k represents the regression coefficient for the *k*th covariate X_{ik}, indicating the effect of X_{ik} on the item easiness parameters. α_j represents the regression coefficient for covariate Z_{ij}, indicating the effect of Z_{ij} on the item discrimination parameters. θ_p represents the random effect for person, which follows a normal distribution, $\theta_p \sim N(0, \sigma^2)$ (Jeon & Rijmen, 2016; Molenberghs & Verbeke, 2004).

When the covariates X_{ik} and Z_{ij} are all item indicator variables, the formula is reduced to

$$logit(P(y_{pi} = 1 \mid \theta_p)) = \alpha_i \theta_p + \beta_i$$

with β_i indicating item easiness and α_j indicating item discrimination (Jeon & Rijmen, 2016). This is the formula of the 2PL model. By using GLNMMs, the effects of different item features can be estimated as regression coefficients for item discrimination and easiness.

Explanatory IRT can provide researchers the necessary information to create a group of new items, which the traditional 2PL or PCM do not provide. In explanatory IRT, a researcher starts by examining the wording of all personality items. Depending on which item feature is of interest, the researcher may use his own judgments or invite SMEs to rate the items. In explanatory IRT analysis, the item feature variables are entered into the model, along with participants' responses to the personality items. The results provide information of how each item feature correlates with item parameter estimates (i.e., α and β). For example, how is item subtlety correlated with item discrimination/easiness estimates? If subtlety demonstrates negative correlation with item discrimination estimates, the researcher knows that subtle items cannot distinguish respondents in general. Then, writing subtle items in future scale development can be avoided. Researchers can also use GLNMMs to examine which item feature is associated with higher/lower item easiness.

GLNMMs have been applied to other fields and have increased the understanding of how person or item features relate to participants' responses. For example, Wilson, De Boeck, and Carstensen (2008) applied GLNMMs to an education test to investigate the effects of two item features, topic areas (i.e. arithmetic, algebra, and geometry), and modeling types (i.e. technical processing, numerical modeling, and abstract modeling) for the German Mathematical Literacy Test. Additionally, De Jong, Pieters, and Fox (2010) applied the GLNMMs to a marketing measure to reduce social desirability in responses to a survey of underreported desires. The GLNMMs provided the flexibility to model responses in the experimental and control groups and allowed for individual-level inferences of the latent trait.

1.3. Item features

To build these models for personality data, we need to include item features that are likely to influence item functioning. In this section, we review the item features likely to influence item discrimination and item easiness parameters of personality items based on previous studies (e.g., Holden, Fekken, & Jackson, 1985; Zickar & Ury, 2002). As discussed previously, the item location parameters in GLNMMs represent item easiness, not item difficulty, and "item location" should be used to represent the *b* parameter in personality items, not "item difficulty" or "easiness." However, researchers usually use 'item location' and 'item difficulty' interchangeably. Thus, in the current study, we still use item easiness to describe item location in personality scales to avoid confusion.

1.3.1. Negative words

Previous studies have found that negatively-worded items might lead to high location and low discrimination parameter estimates (Barker & Ebel, 1982). Negative wording may create an artifact of item wording and change the dimensionality of a scale, and result in lower item discrimination parameters (Greenberger, Chen, Dmitrieva, & Farruggia, 2003; Sliter & Zickar, 2014). Additionally, compared to positive worded items, lower numbers of respondents endorsed negatively worded items (Sliter & Zickar, 2014). People may view negatively worded items as distasteful and in general people are less likely to describe themselves as negative. Thus, we predict that negative wording relates to lower item discrimination and lower item easiness parameters.