Conscientiousness in the workplace: Applying mixture IRT to investigate scalability and predictive validity

Iris J.L. Egberink [a,*], Rob R. Meijer [b], Bernard P. Veldkamp [a]

[a] University of Twente, Faculty of Behavioral Sciences, Department of Research Methodology, Measurement and Data Analysis, PO Box 217, 7500 AE, Enschede, The Netherlands
[b] University of Groningen, Faculty of Behavioral and Social Sciences, Department of Psychometrics and Statistical Techniques, Grote Kruisstraat 2/1, 9712 TS, Groningen, The Netherlands

ABSTRACT

Mixture item response theory (IRT) models have been used to assess multidimensionality of the construct being measured and to detect different response styles for different groups. In this study a mixture version of the graded response model was applied to investigate scalability and predictive validity for a Conscientiousness scale in a career development context ($N = 9283$). A four-class solution yielded the most interpretable results. The classes differed mainly with respect to their scores on the subscales Perfectionism and Concentration. Results showed that Conscientiousness may be qualitatively different for different groups of persons and that the predictive validity of the test scores improved for persons in different classes as compared to fitting a unidimensional IRT model. Implications of this study for personality assessment are discussed.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

Conscientiousness is one of the most important personality traits in the workplace assessed by applied researchers and psychologists because it predicts different types of job performance and various outcomes related to social functioning (e.g., Barrick & Mount, 1991; Barrick, Mount, & Judge, 2001; Dudley, Orvis, Lebiecki, & Cortina, 2006; Roberts, Chernyshenko, Stark, & Goldberg, 2005). Results of several studies show, however, that it is important to pay attention to the lower-order structure of Conscientiousness, in particular to assess predictor–criterion relationships in industrial and organizational (I/O) psychology. For example, Dudley et al. (2006) specified four narrow traits (achievement, order, cautiousness, and dependability) and examined the predictive power of these narrow traits. They found that narrow traits do have incremental validity above and beyond global Conscientiousness, although the incremental validity depended on the particular performance criteria and the occupation in question.

An interesting alternative to obtain more detailed information about the lower-order constructs of Conscientiousness is to consider Conscientiousness scales and subscales from an individual perspective. This would lead to a more person-centered approach instead of a variable-centered approach, which is more common in personality research. In a variable-centered approach the relation between items and subsets of items are the main focus of

interest, whereas in a person-centered approach the main focus is on the person and differences between persons (e.g., differences in the use of the response scale at the individual level). From a psychological assessment point of view, it is most interesting to investigate whether there are individual differences in the way subsets of items are interpreted. This is important in personality assessment because if there are differences, this may affect individual classification and the predictive validity of individual test scores (e.g., Austin, Deary, & Egan, 2006; Weekers & Meijer, 2008).

With the increasing popularity of item response theory (IRT, e.g., Embretson & Reise, 2000) models, techniques have been proposed to analyze individual response patterns and to uncover subgroups that have different probabilities to endorse an item. One such technique is mixture IRT (e.g., Rost, 1990; Rost & Langeheine, 1997). Mixture IRT models combine latent class models and IRT models by identifying groups of individuals (i.e., latent classes, instead of manifest classes, such as gender or age) within a given sample for whom a specific IRT model is applicable. These classes differ from each other with regard to their use of the response scale. Mixture IRT models have been applied in the personality domain for different purposes. Reise and Gomel (1995) suggested that mixture IRT models can be of help to uncover groups of persons who are qualitatively distinct with respect to the probability of endorsing a particular set of items. That is, groups of persons may differ in the way a psychological construct is applicable. Eid and Zickar (2007) used mixture IRT models in the personality domain to uncover faking and response styles. Mixture IRT models have also been applied in other research domains. For example, Eid and Rauber (2000) used these models to detect

measurement invariance in organizational surveys and Muthén and Asparouhov (2006) applied these models to investigate tobacco dependence criteria.

The aim of the present study was threefold: (1) to investigate the number of classes that differ systematically in their response scale usage of a Conscientiousness scale consisting of four lower-order constructs (Perfectionism, Organization, Concentration, and Methodicalness), (2) to explore the characteristics that distinguish the particular classes, and (3) to evaluate the predictive validity of the mixture IRT trait estimates as compared to the estimates under a unidimensional IRT model by relating the estimates to an external criterion measure. Therefore, we first discuss (a) research concerning the structure of Conscientiousness and its predictive validity, (b) the principles of IRT and mixture IRT, and (c) some recent applications of mixture IRT. Second, we apply mixture IRT to a Conscientiousness scale and discuss the differences between the classes with respect to the scalability and the predictive validity of the mixture IRT trait estimates compared to the unidimensional IRT trait estimates. Finally, we reflect on the usefulness of this method for personality assessment.

1.1. The structure and the predictive validity of Conscientiousness

1.1.1. Structure of Conscientiousness

There are different personality questionnaires that take Conscientiousness into account, sometimes consisting of different subscales. Also, the theoretical and empirical underpinnings of these different questionnaires may differ. Therefore, several researchers investigated the lower-order structure of Conscientiousness by analyzing item content and factor analyzing different questionnaires. For example, Saucier and Ostendorf (1999) examined the structure of 500 adjectives of the Big Five. For the Conscientiousness factor, they found four subcomponents (orderliness, decisiveness–consistency, reliability, and industriousness). Peabody and De Raad (2002) combined the results of six studies in different languages to develop a lower-order structure for each Big Five factor. For Conscientiousness, they identified the subcomponents orderliness, work, responsibility, and impulse control as only related to Conscientiousness.

In general, different studies that examined the lower-order structure of Conscientiousness (e.g., MacCann, Duckworth, & Roberts, 2009; Peabody & De Raad, 2002; Perugini & Gallucci, 1997; Roberts, Bogg, Walton, Chernyshenko, & Stark, 2004; Roberts et al., 2005; Saucier & Ostendorf, 1999) identified three common subcomponents, related only to Conscientiousness and not to another Big Five factors: orderliness (i.e., being neat and organized), industriousness (i.e., hard working and being ambitious), and impulse control (i.e., being careful, patient, and cautious). A sometimes identified fourth subcomponent is responsibility, also labeled as reliability and dependability (i.e., being trustworthy, responsible and dependable). However, this subcomponent is often suggested to be a mix of Conscientiousness and Agreeableness (MacCann et al., 2009) or a mix of Conscientiousness and Emotional Stability (Roberts et al., 2005).

In the present study, we used a subscale measuring Perfectionism which is sometimes also considered a mix of Conscientiousness and Neuroticism (e.g., Roberts et al., 2005). Hamachek (1978) distinguished two forms of Perfectionism, a positive form called 'normal perfectionism' (persons who enjoy pursuing their perfectionistic strivings) and a negative form called 'neurotic perfectionism' (persons who suffer from their perfectionistic strivings).

1.1.2. Predictive validity of Conscientiousness

In I/O psychology several researchers noticed that it is important to pay attention to the lower-order structure of the construct of interest when assessing predictor–criterion relationships (see Hough & Ones, 2001; Hough & Oswald, 2000; Paunonen & Ashton, 2001; Paunonen, Haddock, Forsterling, & Keinonen, 2003). In a meta-analysis, Hurtz and Donovan (2000) found an average corrected criterion-related validity between global Conscientiousness and job performance of $r = .22$. Although broad trait measures generally maximize prediction of overall job performance, narrow trait measures maximize the predictive validity of specific criteria. In order to maximize the validity of narrow traits, traits must be selected on the basis of strong a priori linkages to a criterion. Furthermore, narrow traits may help to understand the personality-based causes of individual differences in working behavior (e.g., Hough & Ones, 2001).

Roberts et al. (2005) compared the criterion-related validity of the six subscales they found (industriousness, order, self-control, responsibility, traditionalism, and virtue) to the criterion-related validity of the overall Conscientiousness scale. Results showed that the subscales had a higher criterion-related validity than the Conscientiousness scale in nearly all cases (see also Dudley et al., 2006).

Recently, MacCann et al. (2009) examined the lower-order structure of 18 IPIP scales relating to Conscientiousness (i.e., 117 items). Their analyses resulted in eight subscales: industriousness, perfectionism, tidiness, procrastination refrainment, control, cautiousness, task planning and perseverance. Only control and perseverance were also related to another Big Five factor, namely Agreeableness and Neuroticism, respectively. When examining the criterion-related validity of the broad Conscientiousness factor compared to the validity of the eight different subscales, only perfectionism and industriousness had a significantly higher relationship to the criteria than Conscientiousness. Perfectionism yielded a stronger relationship with Secondary School Admission Test percentiles and industriousness with absence from class.

1.2. Item response theory and mixture item response theory

1.2.1. Item response theory

Item response theory (e.g., Embretson & Reise, 2000) is a collection of statistical models that can be used to analyze items and scales, to create and administer psychological measures, and to measure individuals on psychological constructs. In most IRT models, test responses are assumed to be influenced by a single latent trait, denoted by the Greek letter θ. For dichotomous (true, false) data, the goal of fitting an IRT model is to identify an item response function (IRF) that describes the relation between θ and the probability of item endorsement. In most IRT models, it is assumed that the probability of item endorsement should increase as the trait level increases; thus, IRFs are monotonically increasing functions.

Compared to classical test theory (CTT), IRT has a number of advantages. One advantage is that to judge the quality of an item, one can transform the item's IRF into an item information function, which shows how much psychometric information (a number that represents an item's ability to differentiate among persons) the item provides at each trait level. Different items can provide different amounts of information in different ranges of a given latent trait. Item and scale information are analogous to CTT's item and test reliability. An important difference, however, is that under an IRT framework, information (precision) can vary depending on where an individual falls along the trait range, whereas in CTT, the scale reliability (precision) is assumed to be the same for all individuals, regardless of their raw-score levels. Another advantage of IRT is that, because it is a model-based approach, it is possible to predict a person's answering behavior when confronted with a particular set of questionnaire items. The IRF gives the probability of endorsing an item for each latent trait value.