Psicothema

Psicothema 2014, Vol. 26, No. 3, 372-377
doi: 10.7334/psicothema2013.322

ISSN 0214 - 9915 CODEN PSOTEG

www.psicothema.com

Application of cognitive diagnosis models to competency-based situational judgment tests

Pablo Eduardo García[1], Julio Olea[2] and Jimmy De la Torre[3]
[1] Instituto de Ingeniería del Conocimiento (IIC-UAM), [2] Universidad Autónoma de Madrid and [3] Rutgers, The State University of New Jersey

Abstract

Background: Profiling of jobs in terms of competency requirements has increasingly been applied in many organizational settings. Testing these competencies through situational judgment tests (SJTs) leads to validity problems because it is not usually clear which constructs SJTs measure. The primary purpose of this paper is to evaluate whether the application of cognitive diagnosis models (CDM) to competency-based SJTs can ascertain the underlying competencies measured by the items, and whether these competencies can be estimated precisely. **Method:** The generalized deterministic inputs, noisy "and" gate (G-DINA) model was applied to 26 situational judgment items measuring professional competencies based on the great eight model. These items were applied to 485 employees of a Spanish financial company. The fit of the model to the data and the convergent validity between the estimated competencies and personality dimensions were examined. **Results:** The G-DINA showed a good fit to the data and the estimated competency factors, adapting and coping and interacting and presenting were positively related to emotional stability and extraversion, respectively. **Conclusions:** This work indicates that CDM can be a useful tool when measuring professional competencies through SJTs. CDM can clarify the competencies being measured and provide precise estimates of these competencies.

Keywords: Cognitive diagnosis models (CDM), G-DINA model, situational judgment tests (SJT), great eight model.

Resumen

Aplicación de los modelos de diagnóstico cognitivo a tests de juicio situacional basados en competencias. **Antecedentes:** muchas organizaciones definen sus puestos de trabajo en base a las competencias profesionales que requieren. La medición de tales competencias mediante tests de juicio situacional (TJS) presenta problemas de validez, en tanto no suele estar claro los constructos que miden. El objetivo principal de este estudio es evaluar si la aplicación de los modelos de diagnóstico cognitivo (MDC) a estos tests permite clarificar y estimar de forma precisa las competencias medidas. **Método:** se aplicó el modelo G-DINA (generalized deterministic inputs, noisy "and" gate) a 26 ítems de juicio situacional que medían competencias profesionales fundamentadas en el modelo great eight. Se aplicó el test a 485 trabajadores de una entidad financiera española. Se examinó el ajuste del modelo a los datos, y la validez convergente entre las competencias estimadas y dimensiones de personalidad. **Resultados:** G-DINA mostró un buen ajuste a los datos, y los factores competenciales estimados adaptarse y aguantar, e interactuar y presentar mostraron una relación positiva con estabilidad emocional y extraversión, respectivamente. **Conclusiones:** este trabajo muestra que los MDC pueden ser una herramienta útil para la medición de competencias profesionales a través de TJS, aclarando las competencias que miden y obteniendo estimaciones precisas de las mismas.

Palabras clave: modelos de diagnóstico cognitivo (MDC), modelo G-DINA, tests de juicio situacional (TJS), modelo great eight.

Competency modeling is a useful tool that guides the specification of repertories of behaviors needed for effective performance at work. Many competency models (i.e., defined sets of competencies) have been proposed, most of them pertaining to the managerial area, which aim to reach a judicious equilibrium between the generality and specificity of work demands. One influential model that covers both managerial and non-managerial positions is the great eight model (Bartram, 2005; Kurz & Bartram, 2002). This competency model consists of a three-tier structure. The bottom tier is made up of 112 component competencies (e.g., acting with confidence); the middle tier is made up of 20 competency dimensions (e.g., deciding and initiating action); and the top tier is made up of eight broad competency factors (e.g., leading and deciding), which are usually referred to as the great eight.

This research is focused on the measurement of competencies (middle tier) through situational judgment tests (SJT). SJT present actual work-related situations using various formats (e.g., paper, video), and ask respondents how they would or should deal with those situations, usually having to choose from various response options. These tests have become very popular among industrial and organizational (I/O) psychologists in the last twenty years because they are cheaper than interviews, can be applied in large-scale hiring contexts, and include work-related skills not easily measured by traditional cognitive and personality tests. Several studies have shown the advantages of this type of tests, which include predictive and incremental validity over work performance criteria, incremental validity over measures of

Received: November 25, 2013 • Accepted: April 23, 2014
Corresponding author: Pablo Eduardo García García
Instituto de Ingeniería del Conocimiento (IIC-UAM)
Universidad Autónoma de Madrid
28049 Madrid (Spain)
e-mail: pablo.garcia@icc.uam.es

cognitive ability and personality, reduced adverse impact, better face validity perceptions of applicants, and greater resistance to faking (Lievens, Peeters, & Schollaert, 2008). However, "it would appear that SJT do not strongly relate to any particular construct but are moderately related to many different constructs" (Ployhart & MacKenzie, 2011, p. 244). Validity studies are still needed to clarify which constructs are being measured when applying these tests.

Cognitive diagnosis models

Cognitive diagnosis models (CDM), also called diagnostic classification models (DCM) by other researchers (e.g., Rupp & Templin, 2008), are multidimensional categorical-latent trait models.

Although CDM share several features with other models such as item response theory (IRT) and confirmatory factor analysis (CFA), one of the most important differences is the conceptualization of the latent trait as categorical (usually referred to as *attributes*), rather than continuous.

CDM are formulated to specify the probability of responding to an item in a specific way, given a vector of attributes that indicates which ones a respondent has mastered and not mastered. For existing work, item responses in CDM are categorical. This is one characteristic that differentiates CDM and IRT from CFA, which models the probability distribution of continuous responses. Although item responses can be both dichotomous or polytomous, for didactic reasons we will focus on CDM for dichotomous responses, as in responding correctly or incorrectly to an item, or endorsing or not endorsing a particular statement.

In CDM, the attribute vectors are referred to as attribute patterns or latent classes, and they are denoted by $\boldsymbol{\alpha}_l = (\alpha_{l1} \cdots \alpha_{lK})$, for $l = 1$ to L, where L represent the total number of attribute patterns (i.e., possible combinations of attributes). The k^{th} element of $\boldsymbol{\alpha}_l$ (i.e., α_{lk}) is equal to 1 if the respondents in latent class l have mastered the k^{th} attribute, and 0 if they have not. For example, the attribute vector $\boldsymbol{\alpha}_l$ (11010) means that respondents in latent class l have mastered attributes 1, 2 and 4**,** but not attributes 3 and 5. Theoretically, there are 2^K possible latent classes. For example, with K =5, there would be $L = 2^5 = 32$ possible latent classes, with $\boldsymbol{\alpha}_l$ (11010) being only one of them.

The inputs needed for CDM to be estimated are item responses, and what is known as a *Q-matrix* (Tatsuoka, 1983). As confirmatory models, CDM require specifying which attributes are needed to respond to each item correctly. This is a task that is typically conducted by domain experts. The item-attribute mapping leads to the J (number of items) × K (number of attributes) binary Q-matrix. The element in row j and column k of the Q-matrix, q_{jk}, is equal to 1 if the k^{th} attribute is required to answer item j correctly; otherwise it is equal to zero. Table 1 gives a Q-matrix for five items and five attributes (where the k^{th} attribute is denoted by α_k). Based on the third row of the Q-matrix, attributes 1, 2 and 5 are required to answer item 3 correctly, whereas attributes 3 and 4 are not.

The main output of CDM for each respondent is a vector of estimates indicating the probability that the respondent has mastered each of the attributes. These probability can be converted in dichotomous scores (i.e., mastery or non-mastery) by comparing them to a cutscore (usually .5; de la Torre, Hong, & Deng, 2010; Templin & Henson, 2006).

Several CDM have been proposed in the last few years (Rupp, Templin, & Henson, 2010), and they vary in the way attributes are combined and formalized to estimate the probability of item responses. To synthesize the various CDM, de la Torre (2011) has proposed a general model, the G-DINA (*generalized deterministic inputs, noisy "and" gate*) model, which allows reformulating many of the existing CDM as special cases of this general model.

For each item, the G-DINA model partitions the latent classes into $2^{K_j^*}$ latent groups, where $K_j^* = \sum_{k=1}^{K} q_{jk}$ represents the number of required attributes for item j. For example, item 3 of Table 1 leads to $K_3^* = 3$ and $2^3 = 8$ possible latent groups. We let $\boldsymbol{\alpha}_{lj}^*$ be the reduced attribute vector whose elements are the required attributes for item j (e.g., for item 3, the reduced vector is $\boldsymbol{\alpha}_{l3}^* = (\alpha_{l1}\alpha_{l2}\alpha_{l5})$). For notational convenience but without loss of generality, we can let the first K_j^* attributes be the required attributes for item j. This will aid in the discussion of the properties of the model. In the G-DINA model, the probability that a respondent from latent class l with attribute pattern $\boldsymbol{\alpha}_{lj}^*$ will answer item j correctly is denoted by $P(X_j = 1|\boldsymbol{\alpha}_{lj}^*) = P(\boldsymbol{\alpha}_{lj}^*)$.

The formulation of G-DINA model in its saturated form (i.e., no restrictions are made) is:

$$P\left(\boldsymbol{\alpha}_{lj}^*\right) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\alpha_{lk} + \sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'}\alpha_{lk}\alpha_{lk'} \cdots + \delta_{j12\ldots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{lk} \quad (1)$$

where δ_{j0} is the intercept for item j. It represents the baseline probability (i.e., the probability of a correct or most effective response option when none of the required attributes is mastered); δ_{jk} is the main effect due to α_k. It is the change in the probability of a correct (or most effective) response as a result of mastering a single attribute (i.e., α_k); $\delta_{jkk'}$ is the interaction effect due to α_k and $\alpha_{k'}$ (first-order interaction effect). It is the change in the probability of a correct response due to the mastery of both α_k and $\alpha_{k'}$, and it represents the impact that is over and above the additive impact of the mastery of the same two attributes; and $\delta_{j12\ldots K_j^*}$ is the interaction effect due to $\alpha_1, \ldots, \alpha_{K_j^*}$. It represents the change in the probability of a correct response due to the mastery of all the required attributes that is over and above the impact of the main and lower-order interaction effects.

The G-DINA model parameters can be estimated using marginalized maximum likelihood estimation (MMLE). The algorithm proposed by de la Torre for performing this estimation is largely similar to that described in detail for the DINA model (de la Torre, 2009), and is written in Ox (Doornik, 2003).

Table 1
Example of Q-matrix

	Attribute				
Item	α_1	α_2	α_3	α_4	α_5
1	1	1	1	1	0
2	0	1	0	0	1
3	1	1	0	0	1
4	1	1	0	0	0
5	0	1	1	1	1

The current study presents an application of the G-DINA model to a competency-based SJT. The two main predictions of the work are: a) When applied competency-based SJTs, which are based on the great eight model (Bartram, 2005), the G-DINA model can provide a good model-data fit; and b) Based on the theoretic model predictions, convergent validity evidence between the G-DINA estimates and measures of personality dimensions can be obtained.

Method

Participants

Four hundred eighty five employees of a Spanish financial company were tested in several professional competencies, as part of a developmental process of the company, through computer-based controlled tasks administrations. They were junior employees (less than two years working for the company), aged between 25 and 31 years old and 50.7% of them being women.

Four industrial/organizational psychologists (i.e., three university professors of organizational psychology and a senior consultant at a research and consulting firm) with expertise in competency modeling specified the 0/1 entries of the Q-matrix.

Instruments

Situational judgment test. The employees were tested through *eValue*, a computer-based system for measuring competencies. The system includes inbox exercises, auto-reports, tasks execution tests, and also situational judgment tasks. This system was developed and is currently marketed by the Knowledge Engineering Institute - Instituto de Ingeniería del Conocimiento (IIC; www.iic.uam.es/en/). For the current study, 26 items, which measure seven competencies of Bartram's middle-tier competencies, were chosen from the situational judgment tasks. See Table 2 for these competencies, and the top-tier factors to which they belong. The situational items were structured as follows: (a) applicant must read a critical incident; (b) this is followed by a question about how he/she would act in that situation; and (c) applicant must choose one of three response options. The correct or most effective response was determined by nine experts who participated in the development of these situational tasks. The interjudge agreement of the nine experts for the 26 items was .79.

Personality questionnaire. To examine the validity of the CDM-derived SJT scores, the Big Five personality questionnaire (Aguado, Lucia, Ponte, & Arranz, 2008) administered to the same participants through the eValue system was used. Bartram (2005) showed empirical relations between the top-tier factors of his model and the Big Five dimensions. Table 2 indicates which factors are related to which dimensions.

Table 2
The middle-tier competencies (Bartram, 2005) measured by the items

Competency	Top-tier factor	Big Five dimension
Relating and networking	Interacting and presenting	Extraversion
Persuading and influencing	Interacting and presenting	Extraversion
Presenting and communicating information	Interacting and presenting	Extraversion
Following instructions and procedures	Organizing and executing	Conscientiousness
Adapting and responding to change	Adapting and coping	Emotional stability
Coping with pressure and setbacks	Adapting and coping	Emotional stability
Deciding and initiating action	Leading and deciding	Extraversion

Note: Top-tier factor = the top-tier factor the competency belongs to; Big Five dimension = the personality dimension the top-tier factor is positively related with, according to Bartram (2005)

Procedure and data analysis

Q-matrix specification. In addition to operational definitions of the seven competencies indicated in Table 2, the four experts were also provided with the response option considered most effective for each item and were then asked to decide which competency or competencies were necessary to choose those responses. The task was based on a Delphi method, conducted over three rounds. In Round 1, each expert performed the task individually. In Round 2, the experts were anonymously provided with the decisions of the other experts and were told they could (not should) change their initial specifications. Finally, in Round 3 the four experts met in person, and they discussed in detail their opinions and remaining differences.

Because a total consensus at the end of Round 3 was not mandatory, different Q-matrices were examined to determine the one most appropriate for the data. To accomplish this, the fit indexes of the different Q-matrices were compared. Assuming a CDM can fit the data, Chen, de la Torre, and Zhang (2013) show that Schwarz's (1976) Bayesian information criterion (BIC) can determine the best fitting Q-matrix when used in conjunction with a general CDM such as the G-DINA model. The matrix with the smallest BIC was selected and used in succeeding analyses.

BIC index, model parameters estimated through the MMLE algorithm, as well as the statistics presented onwards are all outputs provided by the code written by de la Torre (2011) in Ox.

Comparison of G-DINA with other simpler models. Given that G-DINA model is a general model in its saturated form (Equation 1), we also examined whether a simpler model can be applied to data without significant loss in statistical fit. Two of these simpler CDM are the DINO (*deterministic input, noisy "or" gate*; Templin & Henson, 2006) and DINA (*deterministic input, noisy "and" gate*; de la Torre, 2009; Haertel, 1989; Junker & Sijtsma, 2001) models. These models can be formulated as special cases of the G-DINA model by imposing restrictions to Equation 1 (see de la Torre, 2011). They represent pure compensatory and non-compensatory models, respectively. For an item, both models can only differentiate between two possible latent groups in determining the probability of a correct response. The DINO model differentiates between those who master at least one of the required attributes from those who do not master any of them (i.e., not mastering one required attribute can be compensated by mastering another of those required); the DINA model differentiates between those who master all the required attributes from those who do not master at least one of them (i.e., the absence of one required attribute cannot be compensated for by the presence of other attributes).

Because the DINO and DINA models are nested in the G-DINA model, the likelihood ratio test (*LR*) was conducted to determine whether using one of these reduced models led to a significant loss of fit (i.e., $LR \neq 0$). *LR* is computed as

$$LR = [-2LL^{\text{reduced model}}] - [-2LL^{\text{generalmodel}}] \quad (2)$$

and it is approximately χ^2 distributed, with degrees of freedom equal to the difference between the numbers of parameters of the saturated and reduced models.

Absolute fit. In addition to the relative fit analyses presented above to select the best fitting Q-matrix and CDM for the current data, an absolute fit analysis was also conducted to determine whether the selected model fit the data adequately (i.e., the best model is not a bad model). Statistics based on the residuals between the observed and predicted Fisher-transformed correlations and between the observed and predicted log-odds ratios of item pairs can be used to statistically test model misfit (Chen et al., 2013). If the evaluated model fits the data, these statistics should be close to zero for all the items. Considering the tests involve a large number of item pairs (26 items would lead to 325 pairs for each test statistic), Chen et al. propose to examine only the residual with the maximum z-score for each statistic. Rejecting any z-score indicates that the model does not fit at least one item pair adequately.

The Fisher-transformed correlations and log-odds ratios residuals were computed as follows:

$$r_{jj'} = \left| Z\left[Corr\left(X_j, X_{j'}\right)\right] - Z\left[Corr\left(\tilde{X}_j, \tilde{X}_{j'}\right)\right]\right|, \text{and} \quad (3)$$

$$l_{jj'} = \left| \log\left(\frac{N_{11}N_{00}}{N_{01}N_{10}}\right) - \log\left(\frac{\tilde{N}_{11}\tilde{N}_{00}}{\tilde{N}_{01}\tilde{N}_{10}}\right)\right| \quad (4)$$

where $j' \neq j$, *Corr* is the Pearson's product-moment correlation, Z is the Fisher transformation, X_j, and $\tilde{X}_j$ are the observed and model-generated data for item j, respectively, and $N_{yy'}$ and $\tilde{N}_{yy'}$ are the number of observed and predicted examinees, respectively, who scored y (0 or 1) on item j, and y' (0 or 1) on item j'.

The approximate standard errors for these statistics, which are needed for obtaining the z-scores, are computed as follows:

$$\text{SE}[rjj'] = [N-3]^{1/2}, \text{and} \quad (5)$$

$$\text{SE}[ljj'] = [\tilde{N}(1/\tilde{N}_{11} + 1/\tilde{N}_{00} + 1/\tilde{N}_{01} + 1/\tilde{N}_{10}) / N]^{1/2} \quad (6)$$

The resulting z-score is assumed to be approximately normally distributed. A Bonferroni correction is also suggested by Chen et al. (2013) so as not to inflate the Type I error due to the multiple comparisons.

Validity analysis. The G-DINA model provides a vector of estimates per respondent representing his/her expected a posteriori (EAP) probabilities of mastering each one of the competencies. Using a cutscore of .5, participants were classified as either mastering or not mastering those competencies.

Once the 485 participants were classified, we examined the relations between these classifications and the Big Five personality dimensions as described by Bartram (2005). Given that Bartram showed that relations exist between the personality dimensions and the eight top-tier factors, only those stated for the factors Adapting-and-coping and Interacting-and-presenting could be properly examined in this study, because all of their aggregated middle-tier competencies had to be estimated (see Table 2). Bartram showed that the top-tier factor Adapting-and-coping is positively related to *emotional stability,* and the top-tier factor Interacting-and-presenting is positively related to *extraversion*. To examine these relations, two *t* tests were conducted to contrast whether people who had mastered all competencies included in the top-tier factor had a significantly higher score in the corresponding personality dimension than people who had not mastered any of those competencies.

Results

Q-matrix Specification

Table 3 shows the experts' decisions for each item across the three rounds. First, it can be observed that Competency 1 (Relating and networking) was almost not selected in the last round despite being one of the seven competencies expected to be measured. Thus, it was eliminated from the list of competencies, and only the six remaining competencies were considered thereafter.

Secondly, the table shows that for those six competencies, a total consensus was not reached at the end of the last round for items 3, 5, 10, 13, 24 and 26. So three Q-matrices were compared: the elements were equal to 1 if all the experts considered a competency necessary (Q-matrix 1), if at least two experts considered it necessary (Q-matrix 2), and if any expert considered it necessary (Q-matrix 3). The G-DINA model estimated for each one of these three Q-matrices led to BIC indices of 15782.75 (Q-matrix 1), 15749.79 (Q-matrix 2) and 15840.37 (Q-matrix 3). Because Q-matrix 2 (see Table 4) had the lowest BIC, it was the Q-matrix used for the subsequent analyses. As can be seen from the table, most items (15) involved three competencies, 6 items involved two competencies, and 5 items involved only one competency.

Comparison of G-DINA with Other Simpler Models

The two χ^2 tests, each one with 102 degrees of freedom, corresponding to the likelihood ratio tests resulting from

Table 3
Experts' task. Competencies that were considered necessary to choose the most effective response option in each item

	Round				Round		
Item	**1**	**2**	**3**	**Item**	**1**	**2**	**3**
1	**4**,5,**6**	**4**,**5**,**6**	**4**,**5**,**6**	14	4,5	4,**5**	**4**,**5**
2	2,3,4,**5**,6	3,4,**5**,**6**	**3**,**5**,**6**	15	**2**,**3**,7	**2**,**3**,**7**	**2**,**3**,**7**
3	2,3,5,6,**7**	**2**,3,**7**	**2**,3[a],**7**	16	**3**	**3**	**3**
4	1,2,3,6,7	**2**,**6**	**2**,**6**	17	**4**,5,6	**4**,6	**4**
5	1,2,3,5,6	1,3,6	**3**,6[b]	18	1,2,3	**2**	**2**
6	**4**,5,6,7	**4**,**5**,6	**4**,**5**,**6**	19	1,**2**,3,6	**2**,3,6	**2**,**3**
7	**2**,3,6	**2**,3,6	**2**	20	**2**,3,6,7	**2**,3,7	**2**,**3**,**7**
8	4,**5**,6,7	4,**5**,6	**4**,**5**,**6**	21	1,4,5,6	1,**4**,5	1[a],**4**,**5**
9	**4**,**5**,**6**	**4**,**5**,**6**	**4**,**5**,**6**	22	1,**4**,5,6	**4**,5,6	**4**,**5**,**6**
10	1,2,**3**	1,2,**3**	**1**,2[a],**3**	23	1,**4**,**5**,6	**4**,**5**,6	**4**,**5**,**6**
11	2,3,4,**5**,7	2,**5**,**7**	**5**,**7**	24	**2**,**3**,5,6,7	**2**,**3**,6,**7**	**2**,**3**,6[b],**7**
12	1,2,3,7	**2**,**3**,7	**2**,**3**,**7**	25	1,**2**,3,7	1,**2**,3,**7**	1[b],**2**,**3**,**7**
13	**4**,5,6,7	**4**,5,6,7	**4**,5[b],**6**,**7**	26	1,**2**,3,5,7	1,**2**,3,7	**1**,**2**,3[a],**7**

Note: Competencies in boldface were considered necessary by the four experts. Competencies 1 = Relating and networking; 2 = Persuading and influencing; 3 = Presenting and Communicating information; 4 = Following instructions and procedures; 5 = Adapting and responding to change; 6 = Coping with pressure and setbacks; 7 = Deciding and initiating action

[a] Two experts considered the competency necessary

[b] One expert considered the competency necessary

comparing the G-DINA model with the DINO (LR = 328.05) and DINA (LR = 319.17) models were both significant (p<.0005). These results indicate that the more parsimonious models led to a significant loss of fit. Consequently, both a pure compensatory and non-compensatory conceptualization of this SJT were deemed to be inadequate. Thus, the DINO and DINA model were discarded, and the G-DINA model was further examined for its adequacy to model the SJT data.

Absolute Fit

The maximum z-scores of residuals between the observed and predicted Fisher-transformed correlation and the observed and predicted log-odds ratios of item pairs were 2.65 and 1.73, respectively. After taking into account the 325 item pairs through the Bonferroni correction, both these statistics were nonsignificant at $\alpha = .01$. Thus, we concluded that the G-DINA model fitted the SJT data adequately.

Validity analysis

Given six competencies, each of the 485 respondents can be theoretically classified in one of the $2^6 = 64$ possible latent classes. Nevertheless, Table 5 shows the seven most prevalent estimated latent classes contained more than the 50% of respondents.

Concerning the relations between the top-tier factors to which the competencies belong and the Big Five personality dimensions (evidence on convergent validity), the t tests for independent groups (with no equality of variance assumption) were significant in the right direction (see means in Table 6). People who mastered the top-tier factor Adapting-and-coping (i.e., those who mastered both Adapting-and-responding-to-change and Coping-with-pressure-and-setbacks) had significantly higher *emotional stability* than people who did not (t = -3.372; p = .001; Cohen's d = 0.57); people who mastered Interacting-and-presenting (i.e., those who mastered both Persuading-and-influencing and Presenting-and-communicating-information) had significantly higher *extraversion* scores than people who did not (t =-2.682; p = .009; Cohen's d = 0.48). It should be noted that in the case of *extraversion*, only two of the original three competencies from the top-tier factor were used because Relating-and-networking was discarded during the Q-matrix specification phase.

Table 4
Q-matrix

	Competency							Competency					
Item	**1**	**2**	**3**	**4**	**5**	**6**	**Item**	**1**	**2**	**3**	**4**	**5**	**6**
1	0	0	1	1	1	0	14	0	0	1	1	0	0
2	0	1	0	1	1	0	15	1	1	0	0	0	1
3	1	1	0	0	0	1	16	0	1	0	0	0	0
4	1	0	0	0	1	0	17	0	0	1	0	0	0
5	0	1	0	0	0	0	18	1	0	0	0	0	0
6	0	0	1	1	1	0	19	1	1	0	0	0	0
7	1	0	0	0	0	0	20	1	1	0	0	0	1
8	0	0	1	1	1	0	21	0	0	1	1	0	0
9	0	0	1	1	1	0	22	0	0	1	1	1	0
10	1	1	0	0	0	0	23	0	0	1	1	1	0
11	0	0	0	1	0	1	24	1	1	0	0	0	1
12	1	1	0	0	0	1	25	1	1	0	0	0	1
13	0	0	1	0	1	1	26	1	1	0	0	0	1

Note: 1 = the competency is required to choose the most effective response option; 0 = the competency is not required to choose the most effective response option. Competencies 1 = Persuading and influencing; 2 = Presenting and Communicating information; 3 = Following instructions and procedures; 4 = Adapting and responding to change; 5 = Coping with pressure and setbacks; 6 = Deciding and initiating action

Table 5
Seven most prevalent estimated latent classes

	Competency					
%	**1**	**2**	**3**	**4**	**5**	**6**
9.46	1	1	0	0	1	1
8.25	1	0	0	0	1	0
8.08	0	1	0	1	0	0
7.93	1	1	1	0	1	1
7.05	1	1	1	1	1	1
6.85	1	0	0	1	1	0
6.09	1	0	1	1	1	0

Note: % of participants in the latent class; 1 = latent class masters the competency; 0 = latent class does not master the competency. Competencies 1 = Persuading and influencing; 2 = Presenting and Communicating information; 3 = Following instructions and procedures; 4 = Adapting and responding to change; 5 = Coping with pressure and setbacks; 6 = Deciding and initiating action

Table 6
Relations between the top-tier competency factors and the Big Five personality dimensions

	Emotional stability			Extraversion	
Adapting and coping	**n**	**Mean (SD)**	**Interacting and presenting**	**n**	**Mean (SD)**
Master	211	69.26 (7.91)	Master	208	56.94 (6.72)
Non master	54	64.59 (9.34)	Non master	47	53.58 (7.99)

Note: Master = $EAP \geq .5$ for all the aggregated competencies; Non master = $EAP < .5$ for all the aggregated competencies; SD = standard deviation

Discussion

This study represents a successful application of CDM to a competency-based SJT. On the one hand, the Q-matrix specification clarified which competencies each of the 26 items measures, providing empirical and statistical evidence of content validity. On the other, the estimated competencies were related with personality variables in the right direction, providing evidence of convergent validity.

This use of CDM for measuring competencies may fill some gaps left unaddressed by traditional approaches to SJT scoring. Computing a single overall score disregards the multidimensional nature of SJT (Christian, Edwards, & Bradley, 2010). In contrast, CDM fully takes into account this multidimensionality, and the meaning of the competency estimates can easily be examined vis-à-vis other measures to clarify their validity, which is one of the most essential matters to be taken into account when developing and evaluating tests (AERA, APA, & NCME, 1999). Nevertheless,

it must be highlighted that the application of CDM involves considering competencies as categorical, a statement with which several authors may not agree.

For future works involving CDM, it would be interesting to evaluate CDM-based SJT predictions against empirical work performance data. It would also be interesting to apply these models to other organizational areas where ratings of competencies are involved. In many assessment centers, judges usually have to perform a two-step evaluation. In the first step, judges have to evaluate how well the participants perform on each exercise. In the second step, judges have to determine the participants' mastery of several competencies based on their performance across multiple exercises. The application of CDM, through the pre-specification of the Q-matrix, would make this second evaluation unnecessary.

Finally, a limitation of the CDM application presented in the current study is that the specification of the competencies measured by each item of the SJT was done after the test was developed. A more optimal approach is to apply these theory-based specifications during the test development itself (de la Torre, Tjoe, Rhoads, & Lam, 2013).

Acknowledgements

This research was supported in part by the UAM-IIC Chair for Psychometric Models and Applications.

References

AERA, APA, & NCME (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Aguado, D., Lucia, B., Ponte, G., & Arranz, V. (2008). Análisis inicial de las propiedades psicométricas del Cuestionario BFCP Internet para la Evaluación de Big-Five [Initial analysis of the psychometric properties of the BFCP Internet questionnaire for the evaluation of Big-Five]. *Revista Electrónica de Metodología Aplicada, 13*(2), 15-30.

Bartram, D. (2005). The Great Eight Competencies: A criterion-centric approach to validation. *Journal of Applied Psychology*, *90*, 1185-1203.

Chen, J., de la Torre, J., & Zhang, Z. (2013). Relative and absolute fit evaluation in Cognitive Diagnosis Modeling. *Journal of Educational Measurement, 50*, 123-140.

Christian, M.S., Edwards, B.D., & Bradley, J.C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology, 63,* 83-117.

de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, *34*, 115-130.

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76*, 179-199.

de la Torre, J., Hong, Y., & Deng, W. (2010). Factors affecting the item parameter estimation and classification accuracy of the DINA model. *Journal of Educational Measurement, 47*, 227-249.

de la Torre, J., Tjoe, H., Rhoads, K., & Lam, D. (2013). Conceptual and theoretical issues in proportional reasoning. *International Journal for Studies in Mathematics Education, 6*, 21-38.

Doornik, J.A. (2003). *Object-Oriented Matrix Programming using Ox (Version 3.1) [Computer software]*. London: Timberlake Consultants Press.

Haertel, E.H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement, 26*, 333-352.

Junker, B.W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with non parametric item response theory. *Applied Psychological Measurement, 25,* 258-272.

Kurz, R., & Bartram, D. (2002). Competency and individual performance: Modeling the world of work. In I.T. Robertson, M. Callinan, & D. Bartram (Eds.), *Organizational effectiveness: The role of psychology* (pp. 227-255). Chichester: Wiley.

Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review, 37*, 426-441.

Ployhart, R.E., & MacKenzie, W.I. (2011). Situational judgment tests: A critical review and agenda for the future. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology (Vol. 2): Selecting and developing members for the organization*. Washington, DC: American Psychological Association.

Rupp. A.A., & Templin, J.L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement*, *6*, 219-262.

Rupp, A.A., Templin J.L., & Henson, R.A. (2010). *Diagnostic measurement. Theory, methods, and applications*. New York, NY: The Guilford Press.

Schwarz, G. (1976). Estimating the dimension of a model. *Annals of Statistics, 6*, 461-464.

Tatsuoka, K.K. (1983). Rule space: An approach for dealing with misconception based on item response theory. *Journal of Education Statistic*, *20*, 345-354.

Templin J.L., & Henson, R.A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods, 11*, 287-305.