THE LATENT CLASS MODEL AS A MEASUREMENT MODEL FOR SITUATIONAL JUDGMENT TESTS

Frank RIJMEN*
ETS, Princeton, USA

In a situational judgment test, it is often debatable what constitutes a correct answer to a situation. There is currently a multitude of scoring procedures. Establishing a measurement model can guide the selection of a scoring rule. It is argued that the latent class model is a good candidate for a measurement model. Two latent class models are applied to the Managing Emotions subtest of the Mayer, Salovey, Caruso Emotional Intelligence Test: a plain-vanilla latent class model, and a second-order latent class model that takes into account the clustering of several possible reactions within each hypothetical scenario of the situational judgment test. The results for both models indicated that there were three subgroups characterised by the degree to which differentiation occurred between possible reactions in terms of perceived effectiveness. Furthermore, the results for the second-order model indicated a moderate cluster effect.

Situational judgment tests consist of several hypothetical but realistic scenarios. After being presented with a scenario, several possible reactions are listed, and the test-taker has to evaluate these responses in terms of appropriateness or the likelihood that he/she would show the response in real life. Situational judgment tests are most often used in an industrial-organisational context, where the scenarios typically originate from critical incidents analysis of a particular job (Weekley & Ployhart, 2006).

More so than is the case for tests of cognitive abilities, it can be a challenging and costly endeavour to identify the "correct" reaction in a given situation. In fact, there are currently many different approaches to establish scoring keys. Bergman, Drasgow, Donovan, Henning, and Juraska (2006) distinguish between empirical scoring, theoretical scoring, and expert-based scoring.

In empirical scoring, the endorsements of listed reactions are correlated with an external criterion for each individual item, and response options are scored such as to maximise that correlation. As Bergman et al. (2006) point

out, different external criteria and differences in practical choices may lead to a multitude of scoring rules. In theoretical scoring, the possible reactions to each situation are scored in line with a certain theory. Of course, the theory may be flawed or there may be alternative theories on what constitutes a proper reaction to the hypothetical situation. In expert-based scoring, scoring keys are based on a consensus between individuals with expertise about the topic. Again, different procedures in reaching consensus and different types of experts may lead to many different scoring rules.

Notwithstanding the variety of possible scoring rules that can be obtained, the three methods have two major characteristics in common. First, all three methods focus on establishing a scoring rule in a rather narrow sense, as opposed to establishing a measurement or psychometric model for situational judgment tests first, and then deriving the scoring rule from the measurement model. An advantage that can be discerned in this respect is that test developers are not overly constrained by psychometric requirements. However, scoring rules naturally follow once a psychometric model is established. In the absence of an underlying psychometric model, any scoring rule may be considered arbitrary. In this respect, the multitude of scoring rules that have been proposed in the literature can be seen as symptomatic for the lack of a psychometric framework for situational judgment tests.

Second, all three methods attempt at arriving at a scoring rule on a situation by situation basis, ignoring additional information that is conveyed by behavioural profiles across situations. Final scores are obtained by aggregating the scores on the individual situations. It may well be, however, that what constitutes an adequate behaviour in one situation depends on how one reacted in other situations. For example, successful leadership may be more a matter of style and being consistent given one's style, rather than having a "correct" way of showing leadership for any given situation. Anecdotal evidence for this can be found in the fact that many people prefer a politician who is of the opposite political spectrum, but consistently so, to a politician who sways according to the popular vote on political issues.

It is hard to judge how substantial the loss of information is due to ignoring response profiles over situations in establishing a scoring rule. This depends on the underlying measurement model, and may vary from one situational judgment test to the other. However, a measurement model is hardly ever specified.

The importance of having an underlying measurement model has been noted by some authors (e.g., Lievens, Peeters, & Schollaert, 2008) when it comes to evaluate the psychometric properties of situational judgment tests. For example, the sometimes rather low internal consistency of scores on situational judgment tests has been attributed to the presumed lack of unidimensionality of the underlying construct that is being measured. That is, the use

of internal consistency has been criticised as a measure for reliability because it is based on a measurement model, i.e., a unidimensional model, which may not be true to the nature of situational judgment tests. In general, having a measurement model can guide not only the development of scoring rules, but also the evaluation of the psychometric properties of the resulting scores.

Towards a psychometric model for situational judgment tests

A psychometric model for an assessment is a formal expression for a theory of the individual differences linked to the assessment. Such a "theory" can be more or less grounded in conceptual knowledge about the construct that is being assessed, and it incorporates certain assumptions about the nature of individual differences. The yardstick to hold a psychometric model against is the degree to which it can account for the associations between a set of manifest variables or items that are indicators of the underlying construct.

For example, in classical test theory, a single continuous (or at least ordered categorical) latent variable is assumed to account for individual differences. The ordered nature of the latent variable implies that all differences between persons on the assessment are quantitative in nature. Insofar as the single latent variable suffices to explain the correlations between the observed variables, the psychometric model holds. In this case, the correlations of the scored items with a total score is a good indication of how well each of the individual items measures the underlying construct, and using the sum, whether or not weighted, of the scores of all items that show a sufficiently large item-total correlation is a defensible practice. Hence, classical test theory leads to relatively simple and transparent scoring rules. The assumption of a single continuous (or ordered) underlying latent variable, however, is a rather strong assumption. In addition, it is taken for granted that what constitutes a correct answer can be determined *a priori*, which is quite a strong presupposition in the light of the apparent lack of consensus on how to score situational judgment tests. Taken together, classical test theory may not be the best choice for a psychometric model. Alternatives should be considered at the very least.

The plain-vanilla latent class model

The psychometric models that are presented in this paper originate in latent class analysis (Goodman, 1974; Haberman, 1974; Lazersfeld & Henry, 1968). Latent class analysis encompasses a whole family of models. In its basic version, the latent class model is a formalisation of the following theory of individual differences. A first tenet is that all associations between a set of categorical response variables are assumed to be accounted for by a categor-

ical unobserved latent class variable. That is, it is assumed that the population consists of a finite number of unobserved subgroups, and that there are no further individual differences within those subgroups. This can be formalised as follows. Let $y_{ij} \in \{1, \ldots, k, \ldots, K\}$ denote the categorical response of person i to item j, with $i = 1, \ldots, I$, and $j = 1, \ldots, J$. Without loss of generality, the number of categories K is assumed to be constant across items. Note that the latent class model can be applied to any set of categorical response variables. There is no need to score response options into correct vs. incorrect answers beforehand. Furthermore, the latent class variable is denoted by z_i, $z_i = 1, \ldots$, $c, \ldots, C$. Each realisation c of z_i is called a latent class. The homogeneity within each latent class is represented by the assumption of conditional independence: Conditional on the latent class variable z_i, the responses are assumed to be conditionally independent,

$$\Pr(y_i|z_i) = \prod_{j=1}^{J} \Pr(y_{ij}|z_i). \tag{1}$$

The marginal probability is obtained as

$$\Pr(y_i) = \sum_{c=1}^{C} \Pr(y_i|z_i = c)\Pr(z_i = c) \tag{2}$$

In its basic version, the latent class model does not impose any restrictions on the conditional response probabilities $\Pr(y_{ij}|z_i)$ and latent class weights $\Pr(z_i)$ other than the obvious restriction that probabilities have to sum to one. Therefore, the model incorporates $(K-1) \times J \times C + C - 1$ parameters:

$$\Pr(y_{ij} = k|z_i = c) = \pi_{kjc}$$

for $k = 1, \ldots, K; j = 1, \ldots J; c = 1, \ldots C;$

$$\text{where } \pi_{Kjc} = 1 - \sum_{k=1}^{K-1} \pi_{kjc} \tag{3a}$$

$$\Pr(z_i = c) = \omega_c \text{ for } c = 1, \ldots C; \text{ where } \omega_c = 1 - \sum_{c=1}^{C-1} \omega_c. \tag{3b}$$

Because no constraints are imposed on this set of response probabilities, the plain-vanilla latent class model is an exploratory model. The lack of order constraints over latent classes implies that the latent classes may differ in a qualitative, rather than quantitative, manner. Latent classes can be interpreted

by inspection of the profiles, across items, of the conditional response probabilities associated with each class. A hypothetical example may make this clearer. Suppose people are asked whether or not they like practicing each of the following sports: fishing, trail running, mountain biking, football, baseball, and chess. For a hypothetical two-class solution, the conditional response probability profiles for liking each of the activities are displayed in Figure 1. The first class is characterised by a high probability of liking fishing, trail running and mountain biking, and low probabilities for liking chess, football and baseball. It could be interpreted as consisting of people who like outdoor activities that are not ball games. The second class is characterised by a preference for fishing and chess compared to the other sports, and could be interpreted as consisting of people who prefer sports that are physically less challenging. To the right of the figure, the prior probabilities of belonging to a latent class are given. In this hypothetical example, a proportion of .80 belongs to the class of outdoor minded people.

   Unlike psychometric models with continuous latent variables, such as the classical test theory and the factor analytic model, the latent class model does not imply a monotonic relation between the position on the latent variable and (sets of) response probabilities. As a consequence, the weighted or raw sum of the scores on individual items is not implied as a scoring rule by the latent class model. Instead, one can compute the posterior probabilities of belonging to each of the classes, given the observed vector of responses $\mathbf{y}_i$, and assign persons to the class for which the posterior probability is the highest.
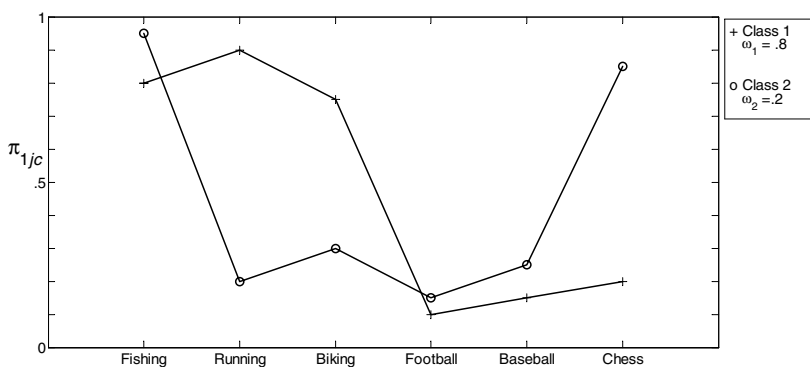

Figure 1
Conditional response profiles for the plain-vanilla latent class model for a hypothetical example

The posterior probabilities are computed through the use of Bayes' theorem:

$$\Pr(z_i|\boldsymbol{y}_i) = \frac{\Pr(\boldsymbol{y}_i|z_i)\Pr(z_i)}{\Pr(\boldsymbol{y}_i)} \tag{4}$$

The latent class model results in a pattern-based scoring rule: each response pattern is mapped onto a latent class.

Conceivably, the measurement theory embodied in the latent class model seems to make it a good candidate for a psychometric model for situational judgment tests. There is no need to define beforehand what constitutes a correct answer, which can be a challenging task for situational judgment tests. Furthermore, latent classes are characterised by a profile of responses across all situations, and hence each latent class can be interpreted as a behavioural style.

Several commercial software packages are available for latent class analysis, with Mplus (Muthén & Muthén, 1998-2010) and Latent GOLD (Vermunt & Magidson, 2005) among the most widely used. The models discussed in this paper were fit with a Matlab toolbox (Rijmen, 2006) for latent variable modelling that can be obtained from the author free of charge.

Application of the plain-vanilla latent class model to the MSCEIT Managing Emotions subtest

The use of the latent class model as a psychometric model for situational judgment tests is illustrated with the Managing Emotions subtest of the Mayer, Salovey, Caruso Emotional Intelligence Test (MSCEIT; Mayer, Salovey, & Caruso, 2002; Mayer, Salovey, Caruso, & Sitarenios, 2003). The Managing Emotions subtest consists of five scenarios of hypothetical situations. For each situation, four possible actions are listed. Each of the four reactions is rated with respect to their effectiveness on a five point Likert scale. More information on the MSCEIT, including some sample items, can be found at http://www.emotionaliq.org/MSCEIT.htm. Ratings were obtained from 716 participants. More information on the sample can be obtained from Roberts, Betancourt, Burrus, Holtzman, Libbrecht, MacCann et al. (2011).

Latent class models with two, three, and four latent classes were fitted on the dichotomised data (1 and 2 *versus* 3, 4, and 5). To reduce the risk of ending up with a local maximum solution, all models were estimated ten times with different random starting values, and the solution with the highest likelihood was chosen.

Table 1 contains the number of parameters, deviance, Akaike information criterion (AIC) and Bayesian information criterion (BIC) for the latent class

models with two, three, and four latent classes. The response profiles for each of the solutions are graphically presented in Figure 2. The AIC hints at a solution with four latent classes, whereas the model with three latent classes is the best model according to the BIC.

In the two-class solution, the first class (represented by the '+' symbol) as a class size of .29, and is characterised by a rather flat profile of conditional response probabilities. So, the persons in this class do not differentiate much with respect to the effectiveness of the listed reactions. Much more differentiation is seen in the response profile for the second class ('o'), consisting of a proportion of .71 of the population. In the second class, there is no single reaction that dominates all other listed reactions in terms of perceived effectiveness for all but the last scenario.

The first class ('+') of the three-class solution has a rather undifferentiated response profile that is very similar to the response profile of the first latent class of the two-class solution. Its class size is .23. The second and third classes both have a conditional response profile that is similar to the response profile of the second class of the two-class solution. The response profile of the second class ('o'), with a class size of .51, is even more differentiated than the response profile of the second class of the two-class solution. The response profile for the third latent class ('*'), which has a class size of .27, is somewhat less differentiated, on the other hand. The third class is an intermediate class, with some of its conditional response probabilities close to the conditional response probabilities of the first class, other close to the conditional response probabilities of the second class, and the remaining probabilities in between those of the first and second class.

For the four-class solution, the response profiles for the first three latent classes closely resemble the response profiles of the latent classes of the three-class solution, with class sizes of .22 for the first class ('+'), .44 for the second class ('o'), and .22 for the third class ('*'). The fourth class ('Δ') has a class size of .11. Its response profile most closely resembles the response profile of the second class, except for two actions for the third scenario, and the last action for the fourth scenario.

Based on the information criteria and the interpretation of the latent class profiles, the three class solution was selected as the starting point for the model that is presented in the next section.

A higher-order latent class model

The plain-vanilla latent class model can be modified and extended in several ways in order to become a measurement model with a tighter fit to the assessment under consideration. For example, background variables of the test-takers (e.g., gender and ethnic background) can be included as covariates

Table 1

Number of parameters, deviance, Akaike information criterion (AIC), and Bayesian information criterion (BIC) for the plain-vanilla latent class model

Classes	Parameters	Deviance	AIC	BIC
2	41	12233	12315	12503
3	62	12084	12208	12492
4	83	12007	12173	12553
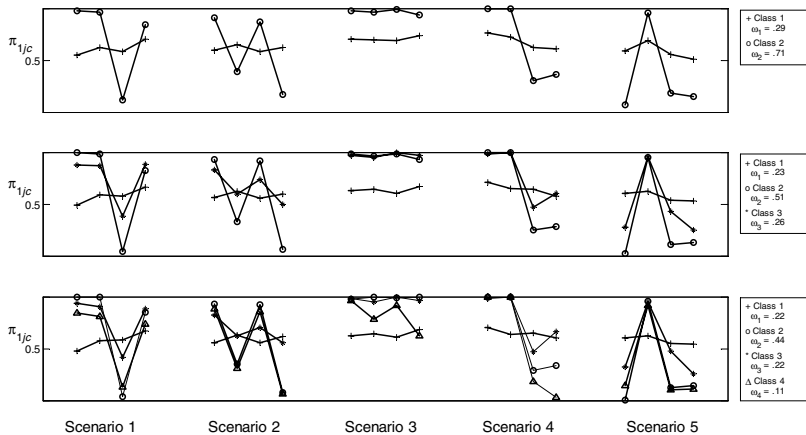
Figure 2

Conditional response probabilities for the plain-vanilla latent class model. Top panel: two class solution. Middle panel: three class solution. Bottom panel: four class solution. Lines connect the four reactions within each of the five scenarios

(Dayton & Macready, 1988; Vermunt, Langeheine, & Böckenholt, 1999). This could be a way to investigate differential item functioning across subgroups. Alternatively, additional continuous latent variables can be included within the latent classes to relax the assumption that there are no individual differences within latent classes (Mislevy & Verhelst, 1990; Rijmen & De Boeck, 2003; Rijmen & De Boeck, 2005; Rijmen, De Boeck, & van der Maas, 2005; Rost, 1990; von Davier, 1994; Yamamoto, 1987). Such a model formalises a theory that states that individual differences can be both qualitative (e.g., leadership styles) and quantitative (e.g., how effective a leader is, given one's style).

In this paper, I will focus on an extension of the latent class model that better takes into account the clustered structure of the data. Specifically, for the situational judgment test for emotional management, four possible actions are clustered within the scenarios describing hypothetical situations. Responses pertaining to the same situation can be expected to be more highly associated

with each other than with responses belonging to different situations. Hence, clusters form an additional source of dependencies, and the presence of cluster effects is a violation of the assumption of conditional independence, the assumption that all responses are independent given the latent variable (Equation 1).

Cluster effects can be taken into account by including an additional latent variable for each hypothetical scenario. A second-order latent class model can then be formulated. At the first level, a separate latent class model is assumed for each scenario. To account for the associations across scenarios, a latent class model is formulated at the second level, which has as indicators the latent class variables of the first level. Such a model is the latent class analogue of a second-order factor model. Denoting the response pattern of a person i for hypothetical situation s, $s = 1, \ldots, S$ by $\mathbf{y}_{is} = (y_{i1s}, \ldots, y_{ijs}, \ldots, y_{iJs})'$, and the corresponding latent variable by z_{is}, Equation 1 becomes

$$\Pr(\mathbf{y}_{is}|z_{is}) = \prod_{j=1}^{J_s} \Pr(y_{ijs}|z_{is}) \tag{5}$$

The situation specific latent variables z_{is} are assumed to be conditional independent given the general latent variable, denoted by x_i,

$$\Pr(z_i|x_i) = \prod_{s=1}^{S} \Pr(z_{is}|x_i) \tag{6}$$

where $z_i = (z_{i1}, \ldots, z_{is}, \ldots, z_{iS})$. With no restrictions on the (conditional) probabilities other than that they should sum to one, the same number of responses J and the same number of latent classes C within each scenario, the same number K of response categories for each response, S scenarios, and G latent classes at the second level, the second-order latent class model has $(K-1) \times J \times S \times C + (C-1) \times S \times G + G - 1$ parameters:

$$\Pr(y_{ijs} = k|z_{is} = c) = \pi_{kjcs}$$

for $k = 1, \ldots, K; j = 1, \ldots J; c = 1, \ldots C; s = 1, \ldots, S;$

$$\text{where } \pi_{kjcs} = 1 - \sum_{k=1}^{K-1} \pi_{kjcs} \tag{7a}$$

$$\Pr(z_{is} = c \mid x_i = g) = \omega_{cgs}$$

for $c = 2, \ldots C; g = 1, \ldots G; s = 1, \ldots S;$

$$\omega_{cgs} = 1 - \sum_{c=1}^{C-1} \omega_{cgs} \tag{7b}$$

$$\Pr(x_i = g) = \tau_c \text{ for } g = 1, \ldots G; \text{ where } \tau_c = 1 - \sum_{g=1}^{G-1} \omega_c \tag{7c}$$

The marginal class weights for each scenario s are obtained as

$$\Pr(z_{is}) = \sum_{g=1}^{G} \Pr(z_{is} \mid x_i = g)\Pr(x_i = g). \tag{8}$$

The idea for a second-order latent class model has been put forward by von Davier and Rost as early as 1996. However, von Davier and Rost did not estimate all the parameters of the model at once, but instead used a two-stage approach, first fitting a separate latent class model for each of a set of item clusters, and then using the assigned class memberships of all persons to the latent classes as the indicator variables for a second latent class analysis. The same approach is followed by Grzywacz, Arcury, Ip, Chapman, Kirk, Bell et al. (2010).

Related latent class models are the multilevel latent class model (Vermunt, 2003) and the hierarchical mixture model for three-way data sets (Vermunt, 2007). In the multilevel model, the clustering occurs at the person mode instead of at the item mode as in the higher-order model. In a three-way dataset, items are fully crossed with another mode (e.g., time) rather than nested as in the higher-order model. All three models nevertheless give rise to similar dependence structures.

Notwithstanding the high-dimensional latent spaces of these models, maximum likelihood estimates can be obtained in an efficient way by exploiting the conditional independence relations between the latent variables. A general method for obtaining an efficient expectation-maximisation algorithm for latent class models through the use of graphical model theory is described by Rijmen, Vansteelandt, and de Boeck (2008). The efficient expectation-maximisation algorithm is implemented in the Matlab toolbox of Rijmen (2006) that was used to fit the models reported in this paper.

Further understanding of the second-order latent class model can be gained by considering two extreme cases. Consider thus, first the case in

which there is no association between the latent variables of the different scenarios, $\Pr(z_{is}|x_i) = \Pr(z_{is})$ for all s. Then, the model reduces to a model in which there is a latent class variable for each scenario, but no second-order latent class variable. When it comes to measurement and scoring, the implications are that the constructs measured in each of the scenarios are unrelated to each other, and therefore scores cannot be aggregated over scenarios in a meaningful fashion.

Second, consider the case in which the number of latent classes is the same at the first and the second level, and there is a perfect one-to-one correspondence between both levels. By implication, the association between the latent classes at the first level is then perfect as well. In this case, the second-order latent class model reduces to the plain-vanilla latent class model. With respect to measurement and scoring, the same construct is assessed in each of the scenarios, and therefore it is meaningful to assign overall scores based on the entire response pattern across scenarios.

Which of the two sketched scenarios is to be preferred based for the situational judgment test for emotional management based on the current data set, is investigated in the next section.

Application of the second-order latent class model to the MSCEIT Managing Emotions subtest

Second-order latent class models with three latent classes at the first level and two to four latent classes at the second level were estimated. The number of classes at the first level was not varied in order to keep the comparison with the results from the plain-vanilla latent class model clear. To reduce the risk of ending up with a local maximum solution, all models were estimated ten times with different random starting values, and the solution with the highest likelihood was chosen. Table 2 contains the number of parameters, deviance, AIC and BIC for each of the models. Both the BIC and the AIC suggested a solution with three latent classes at the second level. In addition, the second-order model with the same number of latent classes at both levels (i.e., three) lends itself for the interpretation outlined in the previous section. Therefore, the three-class model was selected.

The conditional response probabilities, given the latent classes for each of the scenarios at the first level, are presented in Figure 3. Below the response profiles, the marginal latent class weights for the latent classes at the first level are given (see Equation 8) for each scenario.

The conditional response probability profile of the first class for each scenario closely resembles the conditional response profile for that scenario of the first latent class in the plain-vanilla latent class model with three latent classes. The class weights of the first class across scenarios fluctuate around

Table 2

Number of parameters, deviance, Akaike information criterion (AIC), and Bayesian information criterion (BIC) for the second-order latent class model. All models have three latent classes at the first level

Level 2 classes	Parameters	Deviance	AIC	BIC
2	81	12069	12231	12602
3	92	11982	12166	12587
4	103	11966	12172	12643

the class weight of .23 for the first latent class in the plain-vanilla latent class model. Similarly, the second latent class for each scenario is characterised by a response profile that corresponds to the response profile for that scenario of the second latent class in the plain-vanilla latent class model with three latent classes. For the first, second and fourth scenario, the class weights of the second latent class (.55, .51 and .47) do correspond to the class weight of the second latent class in the plain-vanilla latent class model (.51), whereas the class weights of the second class is larger for the two other scenarios (.74 and .69). Finally, the third latent classes for all scenarios but the third one do correspond to the response profiles of those scenarios of the third latent class in the plain-vanilla latent class model with three latent classes. However, the class weight for the third latent class for the third scenario is very small (.03) in the second-order model. Furthermore, in the plain-vanilla latent class model, the response profiles for the second and third class are virtually the same for the third scenario. The third scenario does not seem to differentiate well between those two classes. What seems to be going on for the third scenario is that the second and third latent class of the plain-vanilla latent class model are collapsed onto the second latent class of the second-order model, which explains the large class weight for the second latent class for Scenario 3. Taken together, it can be concluded that the conditional response profiles characterising the latent classes are similar for the plain-vanilla and the second-order model. Hence, the latent classes at the first level of the second-order model correspond to the latent classes in the plain-vanilla latent class model.

The conditional class weights of the latent classes at the first level for each of the scenarios, given the latent class at the second level, are given in Table 3. The last column of the table contains the class weights of the latent classes at the second level.

As explained in the previous section, when there is a perfect association between the latent classes at the first and the second level, the second-order latent class model reduces to the plain-vanilla latent class model. A very strong association of such is present for Scenarios 1, 2 and 4. The conditional probability tables in Table 3 for the other two scenarios do however suggest a more complex relation. While the strong association between the first latent
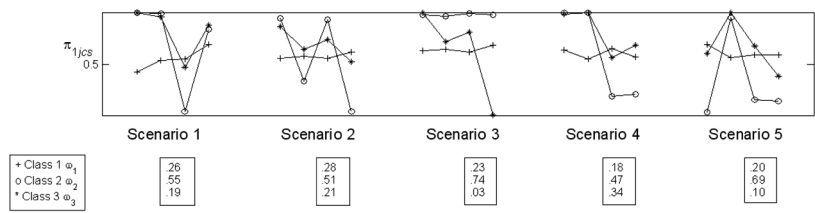
π_{1jcs}
0.5

| Scenario 1 | Scenario 2 | Scenario 3 | Scenario 4 | Scenario 5 |

+ Class 1 ω_1
o Class 2 ω_2
* Class 3 ω_3

.26	.28	.23	.18	.20
.55	.51	.74	.47	.69
.19	.21	.03	.34	.10

Figure 3

Conditional response probabilities for the second-order latent class model. Below
each scenario, the marginal class weights for the latent classes at the first level are
presented

Table 3

Conditional class weights ω_{cgs} of the latent classes at the first level (columns) for
each of the five scenarios, given the latent class at the second level (rows). The last
column contains the class weights of the latent classes at the second level

2nd level	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	τ_g
Class 1 (g = 1)	.93 .00 .07	1.0 .00 .00	1.0 .00 .00	.81 .00 .19	.86 .14 .00	.23
Class 2 (g = 2)	.04 .96 .00	.10 .90 .00	.00 .95 .05	.00 .82 .18	.02 .98 .00	.55
Class 3 (g = 3)	.14 .00 .86	.00 .00 1.0	.03 .97 .00	.00 .04 .96	.00 .50 .50	.22

class at the first and second level holds in these three scenarios as well, the
cross-level associations are weaker for the second and third classes. The con-
ditional probability of pertaining to the second latent class at the first level,
given membership to the second latent class at the second level is high, but a
high conditional probability for belonging to the second class at the first level
given membership to the third latent class at the second level is also observed
for Scenario 3. For Scenario 5, this probability is substantial as well. Analo-
gously, the conditional probabilities for belonging to the third class at the first
level, given that one belongs to the third class at the second level, is only .5
for the fifth scenario and approaches zero for the third scenario.

Altogether, the interpretation of the first two latent classes at the second
level is similar to the interpretation of the first two latent classes of the plain-
vanilla latent class model. The first class consists of persons who do not dif-
ferentiate a lot in terms of the perceived effectiveness of different actions
across a set of hypothetical situations that call for emotional management.
The second class consists of persons who do differentiate. The third class at
the second level of the second-order model exhibits a less stable behaviour
across scenarios. In three scenarios, persons belonging to the third class are
likely to show an intermediate level of differentiation, just as is the case for
the third class of the plain-vanilla latent class model. However, for the fifth

scenario, they are as likely to exhibit a large degree of differentiation, behaving like the second latent class at the second level.

Discussion

The central tenet of this paper is that situational judgment tests are in need of an underlying measurement model. A measurement model is a formal expression of a theory on individual differences related to the assessment. As such, it provides a rationale for the way measures are obtained for the construct that is assessed. For example, if it is assumed in the measurement model that items are more or less independent indicators of a single underlying construct on which persons differ in a quantitative way, the sum score over items can be a good measure of that underlying construct. Without an underlying measurement model, it is hard to justify how the information of several observed variables is combined into a measure of one or more underlying constructs. In addition, it is hard to evaluate the psychometric properties of an assessment without an underlying measurement model because criteria to evaluate the psychometric properties implicitly or explicitly rely on such a measurement model as well. For example, the use of internal consistency measures again implicitly relies on the notion that the items are more or less independent indicators of a single underlying construct.

Latent class models are proposed as candidate psychometric models for situational judgment tests. A first characteristic of latent class models is there is no need to specify beforehand what constitutes a correct answer. Instead, a latent class model can take the raw, not yet scored, categorical responses as input. This characteristic puts latent class models at an advantage, since for many situational judgment tests it may be a challenging and costly endeavour to specify beforehand what constitutes a correct answer. Second, latent class models do not incorporate the assumption that all individual differences can be accounted for by one or more continuous or at least ordered latent variables on which persons differ in a quantitative way. Instead, latent class models assume that the total population consists of homogeneous subgroups that are characterised by response profiles on the observed response variables. These response profiles may reflect qualitative rather than quantitative individual differences, meaning that the conditional response probabilities do not have to be ordered over subgroups. The subgroups are unobserved and represented by one or more categorical latent variables in the model.

A second contribution of this paper is to show how a nested item structure can be accommodated by a second-order latent class model. In this model, a separate categorical latent variable is incorporated for each cluster of items. The association among the latent classes at the first level is accounted for by

a categorical latent variable at the second level, which has as indicators the latent variables at the first level.

Applied to a sample of responses to the MSCEIT Managing Emotions subtest, the response profiles for the latent classes for the plain-vanilla latent class model indicated the presence of a subgroup of respondents that did not differentiate much between the perceived effectiveness of different actions, a subgroup that did differentiate to a large extent, and an intermediate subgroup.

The second-order model further completed this picture by revealing that the third subgroup may exhibit a level of differentiation that depends on the hypothetical situation, differentiating at an intermediate level in some situations, but at a higher level in other situations.

References

Bergman, M.E., Drasgow, F., Donovan, M.A., Henning, J.B., & Juraska, S.E. (2006). Scoring situational judgment tests: Once you get the data, your troubles begin. *International Journal of Selection and Assessment, 14*, 223-235.

Dayton, C.M., & Macready, G.B. (1980). Concomitant-variable latent-class models. *Journal of the American Statistical Association, 83*, 173-178.

Goodman, L.A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika, 61*, 215-231.

Grzywacz, J.G., Arcury, T.A., Ip, E., Chapman, C., Kirk, J.K., Bell, R.A., & Quandt, S.A. (2010). Older Adults' common sense models of diabetes: A multi-method approach illustrating the importance of viewing health beliefs holistically. *Manuscript submitted for publication.*

Haberman, S.J. (1974). Log-linear models for frequency tables derived by indirect observation: Maximum likelihood equations. *The Annals of Statistics, 2*, 911-924.

Lazarsfeld, P.F., & Henry, N.W. (1968). *Latent structure analysis*. Boston: Houghton Mill.

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

Mayer, J.D., Salovey, P., & Caruso, D.R. (2002). *Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) user's manual*. Toronto, Canada: Multi-Health Systems.

Mayer, J.D., Salovey, P., Caruso, D.R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotions, 3*, 97-105.

Mislevy, R.J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika, 55*, 195-215.

Muthén, L.K., & Muthén, B.O. (1998-2010). *Mplus user's guide. Sixth edition.* Los Angeles, CA: Muthén & Muthén.

Rijmen, F. (2006). BNL: A Matlab toolbox for Bayesian networks with logistic regression nodes. *Software manual.* Amsterdam: Free University.

Rijmen, F., & De Boeck, P. (2003). A latent class model for individual differences in the interpretation of conditionals. *Psychological Research, 67*, 219-231.

Rijmen, F., & De Boeck, P. (2005). A relation between a between-item multidimensional IRT model and the mixture-Rasch model. *Psychometrika,70*, 481-496.

Rijmen, F., & De Boeck, P., & van der Maas, H.L.J. (2005). An IRT model with a parameter-driven process for change. *Psychometrika, 70*, 651-669.

Rijmen, F., Vansteelandt, K., & De Boeck, P. (2008). Latent class models for diary method data: Parameter estimation by local computations. *Psychometrika, 73*, 167-182.

Roberts, R.D., Betancourt, A.C., Burrus, J., Holtzman, S., Libbrecht, N., MacCann, C., Matthews, G., Minsky, J., Naemi, B, & Schulze R. (2011). *Multimedia assessment of emotional abilities: Development and validation.* Technical report submitted to ARI.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement, 14*, 271-282.

Vermunt, J.K. (2003). Multilevel latent class models. *Sociological Methodology, 33*, 213-239.

Vermunt, J.K. (2007). A hierarchical mixture model for clustering three-way data sets. *Computational Statistics & Data Analysis, 51*, 5368-5376.

Vermunt, J.K., Langeheine, R., & Böckenholt, U. (1999). Discrete-time discrete-state latent Markov models with time constant and time-varying covariates. *Journal of Educational and Behavioral Statistics, 24*, 178-205.

Vermunt, J.K., & Magidson, J. (2005). Latent GOLD 4.0 user's guide. Belmont Massachusetts: Statistical Innovations Inc.

von Davier, M. (1994). WINMIRA – *A program for analyses with the Rasch model, with the latent class analysis and with the mixed Rasch model.* IPN Software. Kiel, Germany: Institute for Science Education.

von Davier, M., & Rost, J. (1996). Die Erfassung transsituativer Copingstile durch Stimulus-Response Inventare [The assessment of coping styles across situations using stimulus response inventories]. *Diagnostica, 42*, 313-332.

Weekley, J.A., & Ployhart, R.E. (Eds.) (2006). *Situational judgment tests. Theory, measurement, and application.* Mahwah, NJ: Lawrence Erlbaum Ass.

Yamamoto, K. (1987). *A hybrid model for item responses.* Unpublished dissertation, University of Illinois.