



# An analysis of an item-response strategy based on knowledge retrieval

Minjeong Jeon<sup>1</sup> · Paul De Boeck<sup>2</sup>

Published online: 29 June 2018  
© Psychonomic Society, Inc. 2018

## Abstract

In this paper, we hypothesize a type of item-response strategy based on knowledge retrieval. Knowledge retrieval-based item-solving strategy may emerge when: (1) one's regular ability is not utilized, (2) fast response time is not necessarily applied, and (3) the accuracy rate may be higher than the expected level due to chance. We propose to utilize item-response time with a finite-mixture IRT modeling approach to illustrate a potentially knowledge retrieval-based item-solving strategy. The described strategy is illustrated through the utilization of a low-stakes assessment data administered under no time constraint. A simulation study is provided to evaluate the accuracy of the empirical results.

**Keywords** Item-response strategy · Finite-mixture IRT · Response time · Low-stakes assessments · Knowledge retrieval

## Introduction

### Background

Examinees' differential item-response strategies have interested educational researchers for quite some time. For example, Mislevy and Verhelst (1990) proposed a mixed-strategy item-response theory (IRT) model, positing that subjects implement one of a finite number of mutually exclusive item-solution strategies when taking a test. A subject's strategy choice is not directly observed but inferred based on item-response patterns. Yamamoto (1989) proposed a HYBRID model that has a similar finite-mixture structure and can also be used to differentiate examinees' multiple item-solution strategies.

Researchers have paid particular attention to examinees' guessing strategy that often appears during tests administered under time constraints. For instance, Mislevy and Verhelst (1990) applied their model to investigate a random guessing item-response strategy. Yamamoto (1997) proposed an extension of the HYBRID model to identify

examinees who switch their item-solution strategy from a regular to a guessing one. A number of researchers have presented various modeling approaches to capture time-pressured respondents' guessing strategies during speeded tests (e.g., Bolt, Cohen, & Wollack, 2002; Cao & Stokes, 2008; Chang, Tsai, & Hsu, 2014; Wang & Xu, 2015). Unmotivated examinees may also apply a guessing strategy in low-stakes assessments (e.g., Pokropek, 2016; Wise & Kong, 2005; Wise & DeMars, 2006).

We note that guessing is commonly defined in these studies as an item-solving strategy that does not rely on one's ability. Additionally, the following assumptions are implied in these previous studies: (1) guessing is usually a rapid process because a relatively short time is applied compared to a regular ability-based solution strategy; (2) the accuracy of a guessing strategy is within or below the accuracy level expected by chance; (3) only one type of guessing strategy exists, which is differentiable from a regular item-solution strategy.

### Our proposal

Our study aims to examine what one or more secondary latent classes would look like if item-response time is used as covariates for the marginal probabilities of strategy classes. We use the terminology 'secondary class' to contrast it against a class of respondents that supposedly works on the items as intended for the test to measure a certain type of ability. Standard item-response analysis assumes that all examinees will respond to test items

---

✉ Minjeong Jeon  
mjjeon@ucla.edu

<sup>1</sup> Department of Education, University of California, Los Angeles, 3141 Moore Hall, 457 Portola Avenue, Los Angeles CA, 90024, USA

<sup>2</sup> Department of Psychology, Ohio State University, Columbus, OH, USA

based on their ability. In this sense, a class of examinees who may not rely on their ability in solving test items (that we hypothesize and explore in this paper) can be seen as a ‘secondary’ as opposed to a ‘regular’ ability class. It is plausible that response strategies are not only characterized by the patterns of item responses in terms of their accuracy but also based on response time. For instance, a fast guessing strategy would be identified as a class with fast item responses with rather low probabilities of success. This implies that the item response time is an informative covariate for the marginal probability of the guessing strategy class (along with the regular ability-based class if there are only two classes).

It is an empirical question: what would the resulting classes be when response time is used as a covariate to inform the probability of using a strategy? Rapid guessing is not the only possible alternative to the regular, ability-based strategy. Suppose a subgroup of examinees comes to the test with a large body of ready knowledge (obtained from an earlier familiarization with the material) so that for a substantial subset of the items a correct response can be given based on fast retrieval (from the automated knowledge base). Examinees who primarily rely on such a strategy would give fast and correct responses to rather easy items; however, they would spend more time on more difficult items where both knowledge and reasoning processes are needed. We will label this strategy “knowledge retrieval strategy”. In the taxonomy of Bloom (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956), knowledge is considered to be the lowest level of ability or skill development; hence, one may expect subjects at further stages of development to use more complex strategies than just the retrieval of ready knowledge.

The strategy we are referring to, as the basis of a possible secondary class, consists of responding to items on the basis of one’s knowledge even when deeper understanding is available. In principle, it is possible to use this strategy without having a deeper level of understanding, but a deeper understanding does not exclude the possibility that one would respond on the basis of one’s ready knowledge. For example, one may answer a question related to Pythagoras’ theorem on the basis of his or her knowledge of the theorem. The scenario does not exclude the possibility that one has a deeper understanding (thus, can infer and apply the theorem). Similarly, one may know that  $8 \times 7 = 56$  and give an answer on the basis of that knowledge. Even when one is able to (re)construct a correct response on the basis of a deeper understanding, why would one go the long way if a shortcut based on ready knowledge is available? The response can be fast, without much effort, and it would have a high probability of success. However, when an item is more difficult (because it requires reasoning

to make an inference or applying one’s knowledge), trying to retrieve knowledge and make inferences from other bits of knowledge would take additional time and have a lower probability of success.

Independent of the sources of the knowledge, this knowledge may be available at the surface of the knowledge base; hence, this knowledge-based strategy would be fast and have a high probability of giving correct answers. In other cases, the construction of a response or the elimination of incorrect responses is needed based on one’s inferences made from other bits of knowledge. This requires a deeper kind of processing because the correct response is not available at the surface of one’s knowledge base. The terms “surface” and “deep” are used in the literature referring to learning (e.g., Entwistle & Peterson, 2004) and understanding (e.g., Bennet & Bennet, 2008; Jong & Ferguson-Hessler, 1996). Two other related concepts are automated versus controlled processing (e.g., Shiffrin & Schneider, 1977). Retrieving ready knowledge is consistent with automated processing, while controlled processing is intentional, effortful, and requires more time, for example, to make inferences or applications. A knowledge retrieval strategy can in principle be used with surface knowledge as well as with deep knowledge, while it would rely on automated processing rather than on controlled processing.

For test-takers with a rather large ready knowledge base, the knowledge retrieval strategy is a highly efficient strategy because it is successful and does not require much effort. Because one may assume that some examinees using the knowledge retrieval strategy may have some more ready knowledge than other examinees, there may be two or more knowledge-strategy classes that differ with respect to their success rates.

In summary, the knowledge retrieval strategy would show relatively short response time and high probabilities of success, especially for easier items. For more difficult items, a knowledge retrieval strategy would require relatively more time and yet have a low success rate because getting a correct answer requires more than knowledge retrieval but also reasoning processes. This would then lead to relatively lower success rates and longer response time. Hence we expect a positive relationship between item difficulty and the effect of response time on the marginal probability of belonging to the knowledge retrieval class. For easy items, a relatively short response time is an indication for a knowledge retrieval strategy, while for difficult items relatively longer response time may be such an indication. We use the modifying adverb “relatively” here because we will use double centered response time (which will be further explained in section “[Proposed approach](#)”), so that the time intensity of the items and individual differences in item response speed do not play a role in the results.

## Previous studies

Finite-mixture modeling approaches for item-solving strategies have rarely been formulated with item-response time as covariates. The use of response time, however, can throw new light on response strategy modeling, providing a better understanding of response strategies. An alternative way of taking response time information into account is to use a joint mixture model for response accuracy and response time (e.g., Meyer, 2010; Molenaar, Oberski, Vermunt, & De Boeck, 2016; Wang & Xu, 2015). These papers formulated finite-mixture models for item responses, so that responses from the same subject can belong to different latent classes. Models of this type are more flexible in one sense because they can accommodate within-subject variability; but they are less flexible because strong constraints are required to make the models estimable. Specifically, in the Molenaar et al. (2016) model, one class is defined as a class of fast responses and the other class as a class of slow responses. Both of the two classes are differentiated by an item-independent log-time constant. In Meyer (2010), the nature of the two classes is determined a priori. One class is defined as a rapid guessing class and the other class as a solution behavior (ability-based) class. A mixture Rasch model is specified for response accuracy and a mixture log-normal model is specified for response time. The two classes are differentiated by constraining the prior mean (of the response time distribution) for the solution behavior class to be higher than the rapid guessing class. Wang and Xu (2015) also formulated a two-class mixture model where the nature of classes is pre-determined. These authors define one class as a guessing class with an item-specific probability of success and with an item- and person-independent response-time distribution, while setting the other class to follow the hierarchical model for responses and response time, as described by van der Linden (2007), which amounts to a model for response accuracy and a model with latent speed, item time intensity, and discrimination for the response time.

In this study, we will describe a finite-mixture item-response model that also allows for within-subject variation related to the response times, where the variation is the same within the secondary latent class. For example, a knowledge retrieval class would gradually vary from relatively fast responding for successful knowledge retrieval to relatively slow responding for the case where knowledge retrieval does not work well. The secondary class does not need to be a fast or slow class and neither is its nature predefined, such as a rapid guessing class. Although we have speculated on the nature of a possible secondary class, the item success rates and the effects of the response time covariates will in fact be inferred from the data; thus, the nature of a secondary class is an empirical issue. Unfortunately, it

seems impossible to combine the two ideas: (1) within-person mixture models for item responses, and (2) complete openness to the nature of the latent classes.

## Motivating example

Our motivating example is a low-stakes educational assessment, the Major Field Test for the Bachelor's Degree in Business, which is designed for testing undergraduate business students' mastery of concepts, principles, and knowledge. The test items are a mixture of easy and difficult items with a range from knowledge to application.<sup>1</sup> The data included responses from 1000 examinees on 60 multiple-choice psychology items with four response options. A time limit of an hour was given to the examinees to complete the corresponding test section. The average completion time was 42 minutes for the 1000 respondents.

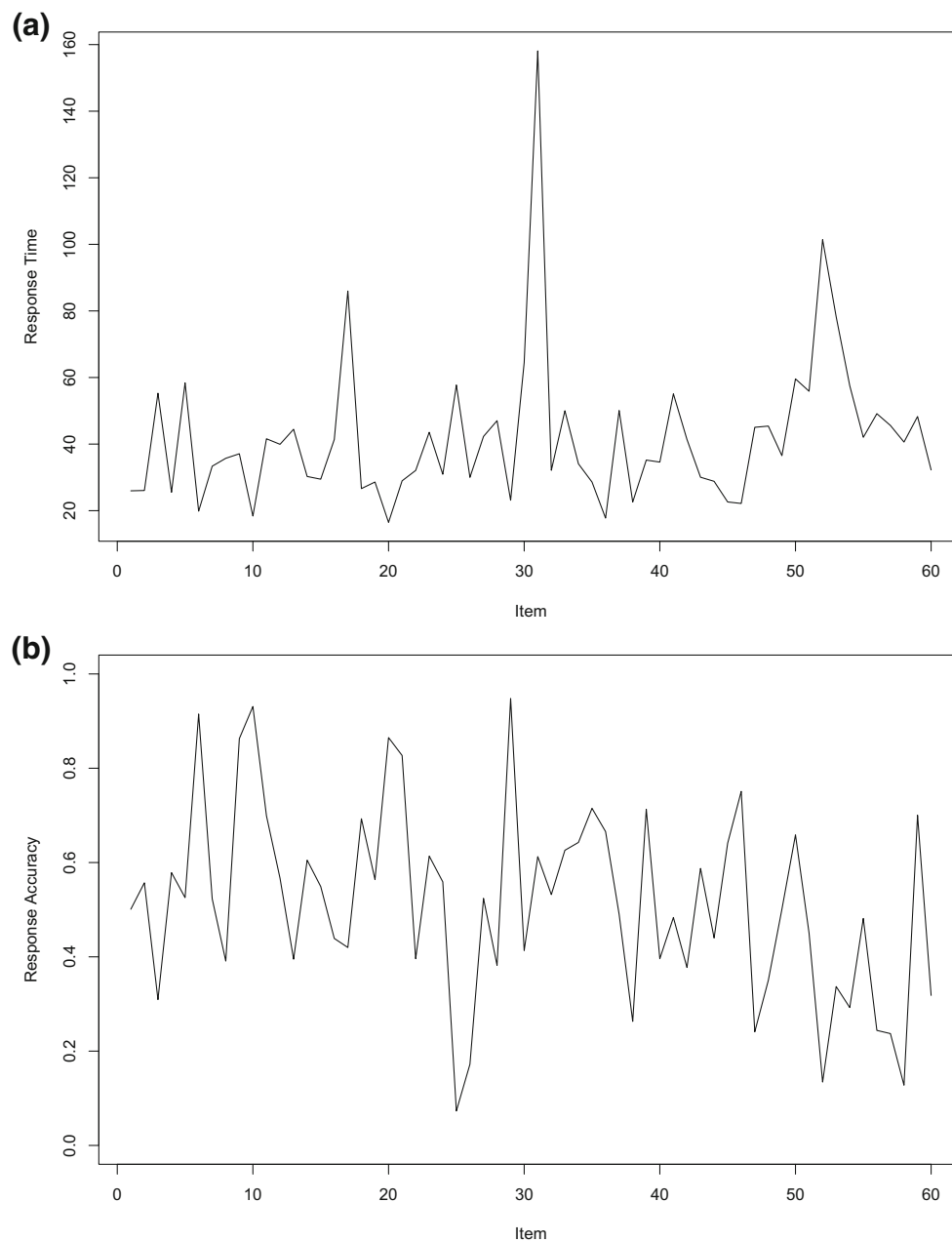
Most examinees gave responses to all the test items (only 0.034% of students provided at least one missing item response). The omitted response rate was very low (less than 0.1%) across all test items (with a median of 0.01%).

The average response time ranged from 16.49 to 158.10 (seconds) across the test items (with a mean of 42.03 and a median of 36.81). In Fig. 1a, the response time distribution showed no distinctive pattern as a function of item location. In addition, we note that fast response time did not seem necessarily applied to solve the end of test items. Figure 2 displays the response time distributions of a set of items selected from the beginning, middle, and end of the test. Figure 2 shows that the response time distributions were alike regardless of the location differences of the items. These distributions did not show clear bimodality that may indicate an existence of a response class with fast responses (or fast guessing). That is, although some of the responses may still be guesses, there are no indications for a substantial proportion of responses being of that kind. This makes sense given that the data were obtained from a low-stakes assessment.

In Fig. 1b, the response accuracy did not show a distinctive pattern across items. The average correct response proportion for all test items was approximately 51.4%. Across all individual items, the average proportions of correct responses ranged from 7 to 95%, while across the 1000 respondents the correct response proportions ranged from 20 to 88%.

These descriptive statistics suggest that there was little evidence of a considerable proportion of random responses

<sup>1</sup>The item contents were unavailable to the authors due to confidentiality issues. Sample items can be found at [https://www.ets.org/Media/Tests/MFT/pdf/mft\\_samp\\_questions\\_business](https://www.ets.org/Media/Tests/MFT/pdf/mft_samp_questions_business). ETS. Copyright © 201x ETS. [www.ets.org](http://www.ets.org). We are grateful to ETS for the permission to use the data.



**Fig. 1** (a) Response time distribution of the 60 test items and (b) Response accuracy distribution of the 60 test items

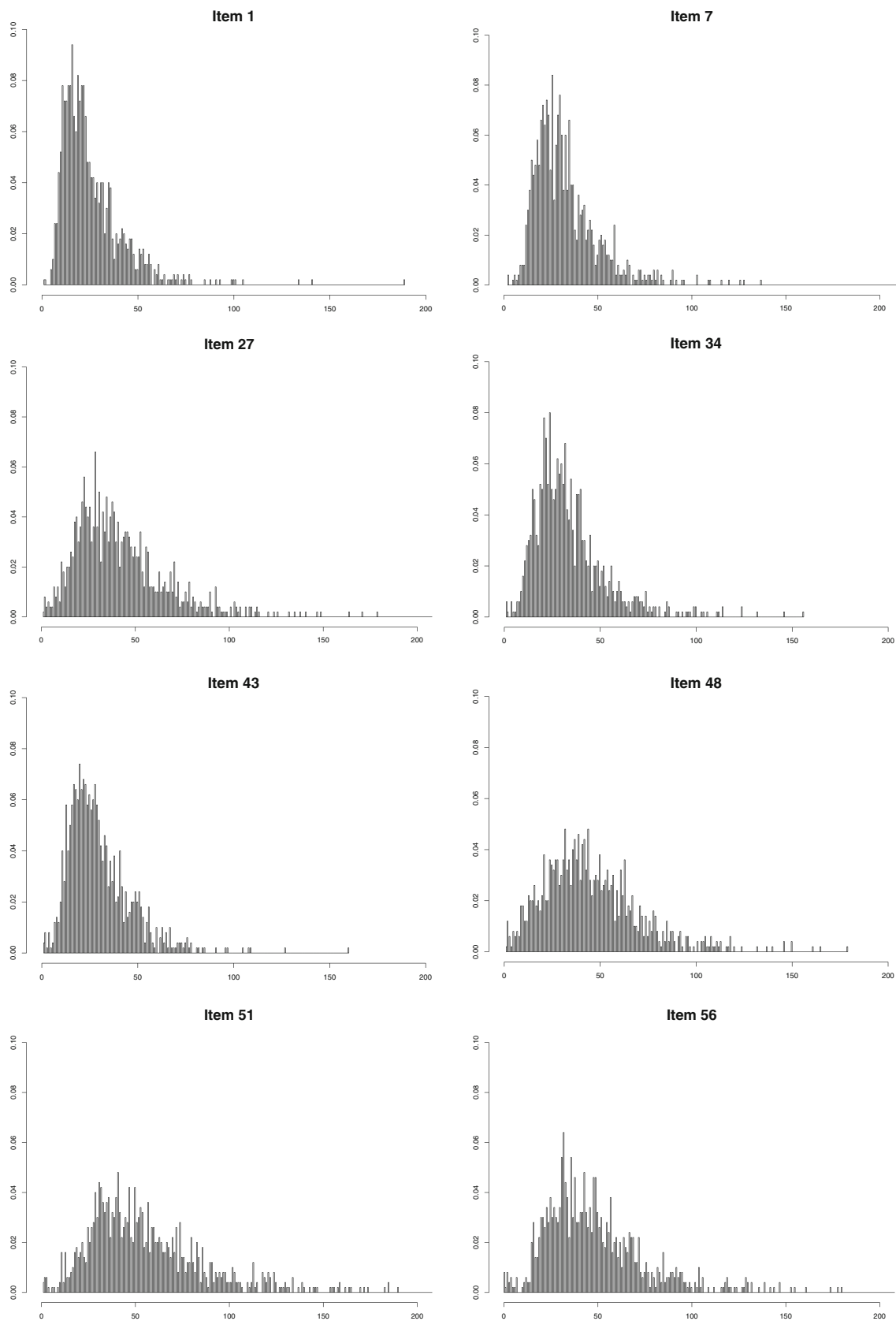
given by time-pressured respondents or by unmotivated respondents in this assessment.

### Proposed approach

Our proposed approach is based on a finite-mixture IRT model that includes a continuous latent variable that represents respondents' regular ability and a discrete latent variable that represents respondents' item solution strategies. We allow for different item response models across different latent classes, where one class follows

a regular IRT model and the secondary classes follow a probabilistic model that lets the data speak about the nature of these classes.

The model includes  $r$  regular classes (ability group or Class  $R_g$ ,  $g = 1, \dots, r$ ) and  $s$  secondary classes (no ability group or Class  $S_g$ ,  $g = 1, \dots, s$ ). We posit a single regular class ( $r = 1$ ) but multiple secondary classes ( $s \geq 1$ ); hence, a total of  $(1 + s)$  latent classes are to be differentiated due to their differential item solving strategies. In this study, we focus on differentiating multiple no-ability-based strategy types, although it is possible that there may also be multiple types of ability-based strategies. One may consider multiple



**Fig. 2** Response time distribution of a selected set of items (Items 1, 7, 27, 34, 43, 48, 51, 56 from the top left to the bottom right, presented by row)

ability-based classes depending on research questions and data conditions.

For the regular class (class  $R$ ),<sup>2</sup> we assume that subjects solve test items in accordance with a regular two-parameter logistic (2PL) model. Specifically, for a binary response  $Y_{ip}$  to item  $i$  ( $i = 1, \dots, I$ ) for subject  $p$  ( $p = 1, \dots, N$ ), the regular, ability group model is specified as follows:

$$P(Y_{ip} = 1 | \theta_p, C_p = R) = \frac{\exp(\alpha_i^{(R)} \theta_p^{(R)} - \beta_i^{(R)})}{1 + \exp(\alpha_i^{(R)} \theta_p^{(R)} - \beta_i^{(R)})}, \quad (1)$$

where  $C_p$  is a categorical latent variable that indicates subject  $p$ 's latent class membership, and  $C_p = R$  represents that subject  $p$  belongs to the regular class. The continuous latent trait variable  $\theta_p^{(R)}$  for subject  $p$  is assumed to be normally distributed, with  $\theta_p^{(R)} \sim N(0, 1)$  (where the trait variance is fixed at 1 for identification). The  $\alpha_i^{(R)}$  and  $\beta_i^{(R)}$  parameters indicate the item discrimination and minus the item intercept parameters, respectively. Equation (1) is specified based on an item factor analysis formulation; hence, the item difficulty parameter can be defined as  $\beta_i^{(R)}/\alpha_i^{(R)}$ . In practice, however, minus the intercept parameter ( $\beta_i^{(R)}$ ) is often interpreted as item difficulty. Hence, for further discussions, we use  $\beta_i^{(R)}$ , instead of  $\beta_i^{(R)}/\alpha_i^{(R)}$ , in this paper.

For the secondary classes (class  $S_g$ ), we postulate that subjects give their item responses without using their regular ability (that one intends to measure with the test). The no-ability based model for the secondary class can be specified as follows:

$$P(Y_{ip} = 1 | C_p = S_g) = \frac{\exp(-\delta_i^{(S_g)})}{1 + \exp(-\delta_i^{(S_g)})}. \quad (2)$$

Note that person latent trait is not incorporated in the probability function in Eq. 2. This means that for subject  $p$  who belongs to the secondary class  $S_g$ , his/her probability of success for item  $i$  is a function of the  $-\delta_i^{(S_g)}$  parameter, where  $\delta_i^{(S_g)}$  is minus the item intercept parameter (representing item difficulty) and  $-\delta_i^{(S_g)}$  indicates the logit of the probability of correctly solving item  $i$  for subjects in class  $S_g$ .

We further postulate that examinees' item response time can be indicative of their strategy choice. Hence, we utilize item response time to predict subjects' probabilities of belonging to class  $S_g$ . We log-transform the response time to achieve approximate normality of the response time distribution (van der Linden, 2007); we then mean-center the log response time both within persons and within

items so that within-person and within-item differences in the log response time can be adjusted. For within-person centering, the mean of the log response times across all items was computed per person ( $\overline{RT}_{.p}^{WP}$ ). For within-item centering, the mean of the log response times across all respondents was computed per item ( $\overline{RT}_i^{WI}$ ). The double-centered response time ( $RT_{ip}^{DC}$ ) was computed as  $RT_{ip}^{DC} = RT_{ip} - \overline{RT}_{.p}^{WP} - \overline{RT}_i^{WI} + \overline{RT}_{..}$ . The main reason for applying this double-centering (for both within-person and within-item centering) to response time data is that when raw response times (log-transformed) are used as covariates, heterogeneity in respondents' speed as well as heterogeneity in item time intensity may confound the estimate results. This was the main criticism on existing modeling approaches that utilize raw response times as covariates (e.g., van der Linden, 2009). As a method to address this confounding issue, van der Linden (2007) and van der Linden (2009) proposed a hierarchical model that utilizes a factor analytic model for response time (that directly models a speed latent variable and item time intensity parameters).

Accordingly, a multinomial logit model is specified (with the ability group as the reference class) for the double-centered log response time ( $RT_{ip}^{DC}$ ) as follows:

$$P(C_p = S_g) = \frac{\exp(\gamma_0^{(S_g)} + \sum_i^I \gamma_i^{(S_g)} RT_{ip}^{DC})}{\sum_{u=1}^{1+s} \exp(\gamma_0^{(u)} + \sum_i^I \gamma_i^{(u)} RT_{ip}^{DC})}, \quad (3)$$

where  $u = R, S_1, S_2, \dots, S_s$  (with  $u = R$  as the reference group),  $\gamma_0^{(S_g)}$  and  $\gamma_i^{(S_g)}$  represent the intercept and the regression coefficient of the log response time for item  $i$  ( $i = 1, \dots, I$ ) for class  $S_g$ . The intercept  $\gamma_0^{(S_g)}$  is fixed at 0 but all  $\gamma_i^{(S_g)}$  parameters ( $i = 1, \dots, I$ ) are freely estimated.<sup>3</sup> The  $\gamma_i^{(S_g)}$  parameters indicate whether or not the log response time for the corresponding test items helps subject  $p$  to be classified into secondary class  $S_g$ . If  $\gamma_i^{(S_g)}$  is positive and significant, it means that spending more time on item  $i$  is more likely to help the subject to belong in the secondary class. On the other hand, a negative  $\gamma_i^{(S_g)}$  value indicates spending more time on item  $i$  is less likely to help the subject to belong in the secondary class. Hence, the estimated  $\gamma_i^{(S_g)}$  patterns can be used to appreciate the nature of the secondary respondent group  $S_g$ . For instance, if  $\gamma_i^{(S_g)} < 0$  and significant for all  $i$ , Class  $S_g$  may be described as a rapid guessing group. However, if  $\gamma_i^{(S_g)} > 0$  for some items, this means that Class  $S_g$  cannot simply

<sup>2</sup>Since we assume that there is one ability-based class, we omit the subscript of  $R_g$  for the sake of simplicity. That is,  $R_g = R_1 = R$  in the current paper.

<sup>3</sup>When the intercept was freely estimated, the value was not significantly different from 0 (also the log-likelihood was identical to the first decimal point); hence, we fixed the intercept to 0 for computational ease.

be defined as a rapid guessing class, because taking a longer time on test items also contributes to generating the secondary group of respondents.

The marginal probability model can be specified with respect to subjects' all potential class memberships:

$$\Pr(y_{ip} = 1|\theta_p) = \Pr(y_{ip} = 1|\theta_p^{(R)}, C_p = R)P(C_p = R) + \sum_{g=1}^s \Pr(y_{ip} = 1|C_p = S_g)P(C_p = S_g), \quad (4)$$

where  $P(C_p = R) = 1 - \sum_{g=1}^s P(C_p = S_g)$  is the probability of belonging to the regular class and the probability of belonging to the secondary class  $P(C_p = S_g)$  is described in Eq. 3.

## Application

We applied the proposed approach to the motivating data example described in section “[Motivating example](#)”. To proceed with the data analysis, we log-transformed the raw response time such that the distribution became close to normal. The resulting log response time distribution had a median/mean of 3.49 and a maximum of 6.86. We additionally applied within-person and within-item mean-centering to the log-transformed response time as discussed in section “[Proposed approach](#)”.

To estimate the proposed model, Mplus was used with full information maximum likelihood estimation with 15 quadrature points. An example Mplus code for the estimated model is provided in the [Appendix](#). To compare the fit of models, we considered two likelihood-based statistics, AIC and sample size adjusted BIC (SABIC). The SABIC places a penalty for adding parameters based on sample size. We used the SABIC rather than the BIC following recent simulation study results that recommend applying the SABIC for model comparisons (e.g., Enders & Tofghi, 2008; Tofghi & Enders, 2007; Yang, 2006).

## Preliminary analysis

We first applied regular finite-mixture 2PL IRT models to the data in order to see whether the data consist of multiple sub-populations with different distributional characteristics (with a 2PL IRT model in each of the classes and no response time covariates). We fit two-class and three-class models (all with ability-based classes) and found that both models were empirically under-identified.<sup>4</sup> This result indicates that there was insufficient evidence for the existence of multiple sub-classes of examinees who

<sup>4</sup>Both models generated a singular information matrix error.

**Table 1** Goodness-of-fit statistics of the estimated models

Model	log-likelihood	p	AIC	SABIC
Preliminary				
Model A1	−34955.319	120	70150.637	70358.441
Model A2a	−35549.760	120	71339.521	71547.324
Model A2b	−34966.241	180	70292.483	70604.189
Proposed				
Model B2	−34677.615	240	69835.231	70250.839
Model B3*	−34640.015	241	69762.031	70179.370
Model B4	−34268.662	480	69497.323	70328.539

Model A1: One-class model; with no RT covariates. Model A2a: Two-class model (with one ability class and one secondary (no-ability) class); with no RT covariates;  $\delta_i^{(S_1)} = 0.25$ . Model A2b: Two-class model (with one ability class and one secondary (no-ability) class); with no RT covariates;  $\delta_i^{(S_1)}$  are freely estimated. Model B2: Two-class RT-based model (with one ability class and one secondary (no-ability) class); with RT covariates. Model B3\*: Constrained three-class RT-based model (with one ability class and two secondary (no-ability) classes); with RT covariates; constrained such that  $\delta_i^{(S_1)} = \delta_i^{(S_2)} + c_1$ . Model B4: Four-class RT-based model (with one ability class and three secondary (no-ability) classes); with RT covariates.  $p$  is the number of parameters and SABIC indicates sample-size adjusted BIC

qualitatively differed in their ability distributions. However, this does not rule out the possibility that there might be additional latent classes based on a strategy that is not reflected in a 2PL model.

In addition, previous studies suggest that unmotivated respondents may apply a guessing strategy during low-stakes assessments (Wise & Kong, 2005; Wise & Demars, 2006). Therefore, we compared a single-class 2PL IRT model (Model 1) with two types of two-class models that include a regular item solution class and a secondary guessing (no ability) class. In one model (Model A2a), the probability of success for the guessing group was fixed at 0.25 (the reciprocal of the number of response options); in another model (Model A2b),<sup>5</sup> the probability of success was not fixed but freely estimated. The former, constrained method is equivalent to Mislevy and Verhelst (1990)'s mixed-strategy model.

The goodness-of-fit of the two mixture models, however, was not better than the regular, single-class 2PL model (see Table 1 for the fit statistics). On the one hand, this result seems to indicate that with the current data it was hard to differentiate unmotivated respondents who gave random responses to the test items. On the other hand, this result also

<sup>5</sup>The number (e.g., 1, 2) appeared in model names (e.g., Model 1, Model A2a) indicates the number of latent classes presumed in the corresponding models.

suggests that we would need to apply a somewhat different analytic approach if we wish to investigate examinees' potentially different item solution strategies that might have been applied to these non-speeded, low-stakes test items.

### Results: Two-class model with response times as covariates

We now apply our proposed approach to the data. We considered a two-class model and incorporated response time information to predict one's classification into the secondary class, as discussed in section "Proposed approach". To differentiate our proposed approach from the two-class models fitted in the preliminary analysis (Models A2a and A2b discussed in section "Preliminary analysis"), we refer to our model as a two-class RT-based model (Model B2).

Results showed that the goodness-of-fit of the two-class RT-based model (Model B2) was improved compared to the fit of the two two-class null models (Models A2a and A2b) as well as the regular 2PL model (Model A1). See Table 1 for details. This means that a secondary strategy class helps to understand the data. Specifically, 37.2% of the examinees were identified within the secondary class, while 62.8% of the examinees were classified into the regular, ability-based class.

The response time information for 17 items was useful to identify the secondary class (statistically significant at the 5% level). The estimated regression coefficients ( $\gamma_i^{(S_1)}$ ) of the response time variable ranged from -2.02 to 1.90 for the 17 items (a mean of 0.05 and SD of 1.2). The fact that the  $\gamma_i^{(S_1)}$  estimates included both positive and negative signs indicates that fast response time was not always indicative of the no-ability based strategy. This means that the identified item response strategy cannot simply be described as fast guessing.

To measure how the impact of the item-response time on person classification was related to the item difficulty, we plotted the  $\gamma_i^{(S_1)}$  parameter estimates (regression coefficients for response time) with the  $\delta_i^{(S_1)}$  parameter estimates (minus item intercept for the secondary class). See Fig. 3a for the result (all parameter estimates and their standard errors are provided in Tables 2 and 3).

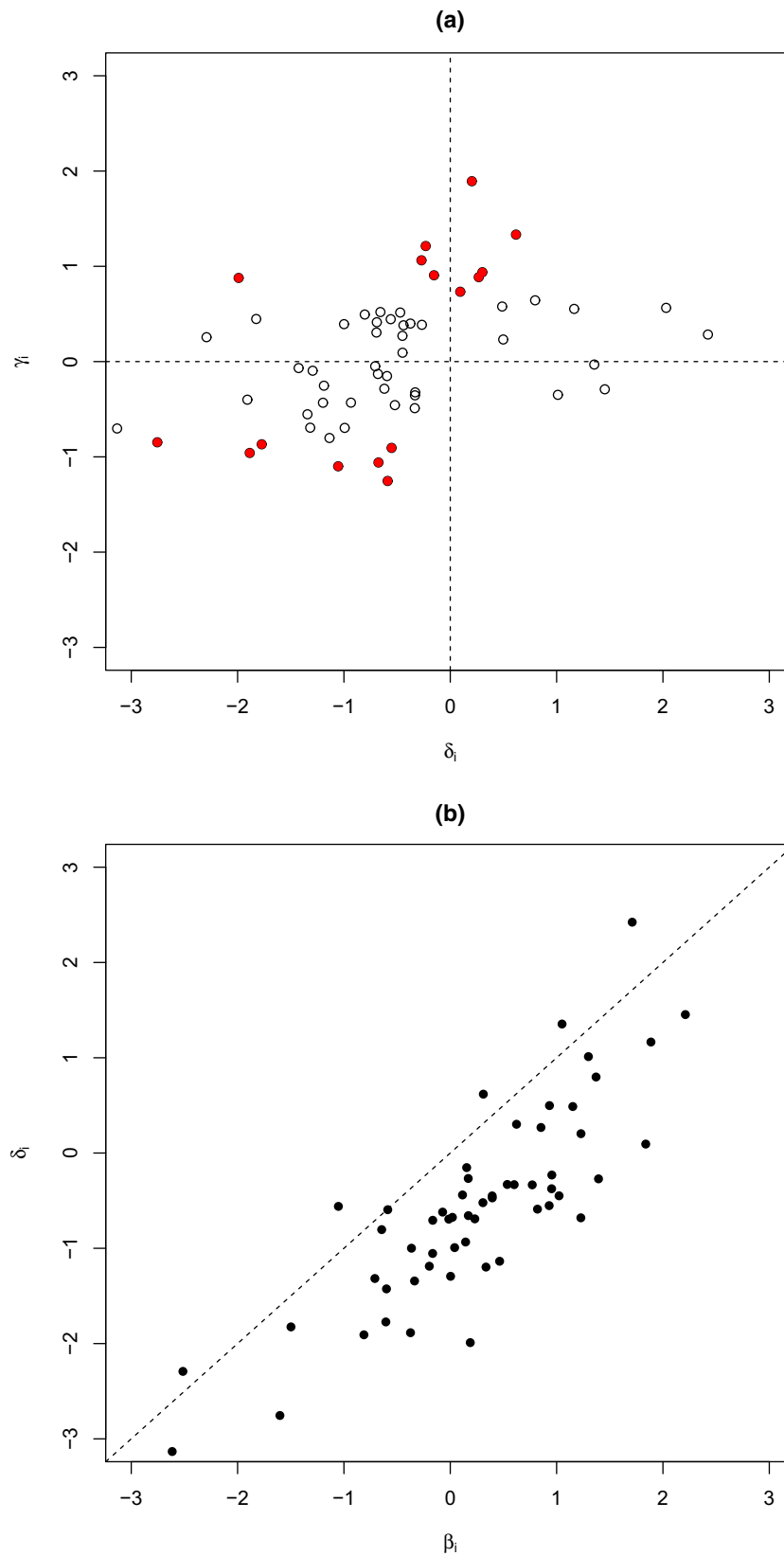
The  $\gamma_i^{(S_1)}$  coefficients tended to be positively associated with the  $\delta_i^{(S_1)}$  parameter estimates (with a correlation of 0.49 for all items and 0.71 for the significant  $\gamma_i^{(S_1)}$  items). When  $\gamma_i^{(S_1)} < 0$ , the  $\delta_i^{(S_1)}$  estimates were often smaller than 0 (bottom left corner). This means that for relatively easy items, spending less time on an item helped the respondents to be grouped into the secondary class. When  $\gamma_i^{(S_1)} > 0$ , the  $\delta_i^{(S_1)}$  estimates were often larger than 0 (top right corner), indicating that for relatively

difficult items, spending more time on an item helped the respondents to be assigned into the secondary class. That is, fast responses tended to result in higher probabilities of correct answers, while slow responses led to lower success rates. This type of secondary class with relatively fast responses for rather easy items and relatively slower responses for rather difficult items cannot be attributed to the well-known relationship between easiness and short response times because the item differences in response time are eliminated by double-centering. Similarly, the secondary class cannot be interpreted as a general speed related class because individual differences in response time are also eliminated. A possible interpretation is that the secondary class represents a mixture of strategies: knowledge retrieval when it works well and less successful strategies if knowledge retrieval does not work. We will denote the secondary class as a knowledge retrieval class although it is a hypothetical interpretation and although alternative strategies may have been used if knowledge retrieval does not work.

We then compared the average raw response time for the no-ability class with the response time for the regular class. The most likely latent class for individual respondents was obtained using maximal a posteriori estimation. The item-level average response time for the no-ability group ranged from 14.3 to 186 (with a median 36.5), while for the regular class it ranged from 17.0 to 141 (with a median of 36). The person-level average response time for the no-ability group ranged from 18.3 to 59.5 (with a median 36.5), while for the regular class it ranged from 2.2 to 73.1 (with a median of 41.4). That is, the respondents in the secondary class did not show a consistent pattern (i.e., consistently fast or slow response time) compared to those examinees in the regular class.

Next, we assessed the relationship between minus the item intercept parameter estimates (item difficulties) for the regular class ( $\beta_i^{(R)}$ ) and for the secondary class ( $\delta_i^{(S_1)}$ ). Figure 3b shows the results. The two parameter estimates were positively correlated (with a correlation of 0.867) although the  $\delta_i^{(S_1)}$  parameter estimates (for class 2) tended to be smaller than the  $\beta_i^{(R)}$  parameter estimates. The average difference between the two parameter estimates ( $\delta_i^{(S_1)} - \beta_i^{(R)}$ ) was -0.738 (SE = 0.031). This means that the test items tended to be easier for those respondents who applied the no-ability strategy. In other words, the secondary class respondents were generally more successful in solving the test items than the regular class respondents. This is a highly intriguing result because if the secondary class respondents had applied a guessing strategy, their performance would have been poorer compared to a standard ability-based strategy (Wise & Kong, 2005). Furthermore, for most items, the probability of a correct response for the secondary class





**Fig. 3** (a) Scatter plot of  $\delta_i^{(S1)}$  and  $\gamma_i^{(S1)}$  (filled red dots for the items with significant  $\gamma_i^{(S1)}$  estimates) and (b) scatter plot of  $\beta_i^{(R)}$  and  $\delta_i^{(S1)}$  from the two-class RT-based model (Model B2)

**Table 2** Item parameter estimates (for Items 1 to 30) from the two-class RT-based model (Model B2, with an ability class and a secondary class)

Item	$\alpha_i^{(R)}$		$\beta_i^{(R)}$		$\delta_i^{(S_1)}$		$\gamma_i^{(S_1)}$	
	EST	SE	EST	SE	EST	SE	EST	SE
1	0.33	0.15	1.30	0.11	1.01	0.13	-0.35	0.36
2	0.28	0.17	-1.05	0.11	-0.56	0.13	0.44	0.47
3	0.29	0.26	2.21	0.16	1.45	0.16	-0.29	0.37
4	-0.10	0.15	0.62	0.09	0.30	0.13	0.94	0.40
5	-0.22	0.22	1.89	0.14	1.16	0.16	0.55	0.40
6	0.67	0.34	3.19	0.25	2.03	0.19	0.56	0.31
7	-0.22	0.15	1.37	0.11	0.80	0.15	0.64	0.38
8	0.10	0.15	0.93	0.10	0.50	0.12	0.23	0.35
9	0.07	0.14	1.15	0.10	0.49	0.13	0.58	0.45
10	0.26	0.16	0.85	0.10	0.27	0.12	0.89	0.41
11	0.11	0.16	1.23	0.11	0.20	0.12	1.89	0.53
12	0.08	0.21	1.84	0.14	0.09	0.14	0.73	0.36
13	0.21	0.17	1.39	0.12	-0.27	0.15	1.06	0.41
14	0.15	0.15	0.95	0.10	-0.23	0.12	1.21	0.48
15	0.12	0.17	0.95	0.11	-0.38	0.14	0.40	0.39
16	0.35	0.16	1.02	0.11	-0.45	0.13	0.09	0.38
17	0.15	0.17	0.93	0.11	-0.55	0.15	-0.91	0.40
18	0.27	0.16	0.60	0.10	-0.33	0.12	-0.35	0.35
19	0.46	0.15	0.77	0.10	-0.33	0.13	-0.49	0.37
20	0.30	0.20	1.23	0.13	-0.68	0.14	-0.13	0.46
21	0.37	0.14	0.54	0.09	-0.33	0.14	-0.32	0.39
22	0.60	0.18	0.82	0.11	-0.59	0.13	-1.25	0.42
23	0.17	0.13	0.15	0.09	-0.15	0.12	0.91	0.40
24	0.28	0.15	0.39	0.09	-0.45	0.13	0.27	0.57
25	0.38	0.14	0.40	0.10	-0.47	0.13	0.52	0.31
26	0.41	0.15	0.31	0.09	-0.52	0.14	-0.46	0.31
27	0.71	0.17	0.17	0.11	-0.27	0.12	0.39	0.45
28	0.22	0.14	0.46	0.10	-1.14	0.18	-0.80	0.47
29	0.15	0.13	0.23	0.09	-0.69	0.15	0.41	0.33
30	0.46	0.16	0.12	0.09	-0.44	0.13	0.38	0.42

$\alpha_i^{(R)}$  and  $\beta_i^{(R)}$  are the item discrimination and minus intercept parameter estimates for the ability class and  $\delta_i^{(S_1)}$  are minus the intercept parameter estimates from the secondary (non-ability) class.  $\gamma_i^{(S_1)}$  are the regression coefficients of response time on the probability of being classified into class  $S_1$

respondents was generally higher than .25 (the accuracy expected by chance). See Fig. 4 for the result.

Taken together, these results show that respondents identified in the secondary class might not be simply a group of random *guessers*. These individuals might be better described as *knowers* (considering their response time and the item response success rates). It may be reasonable to speculate that knowers belong to a knowledge retrieval class. The knowledge retrieval strategy is different from using pre-knowledge on the items, stemming from item over-exposure or compromise (e.g., Shu, Henson, & Luecht, 2013), in that the test was a low-stakes assessment of little consequence.

### Results: Three-class model with response times as covariates

In order to find out whether the secondary class from the two-class RT-based model could further be differentiated, we applied a three-class RT-based model with one regular class and two secondary (no-ability) classes (Model B3). The three-class model was, however, empirically unidentified;<sup>6</sup> hence, we modified the model by imposing the assumption that there were systematic differences in

<sup>6</sup>The information matrix was singular unless some parameter constraints were imposed.

**Table 3** Item parameter estimates (for Items 31 to 60) from the two-class RT-based model (Model B2, with an ability class and a secondary class)

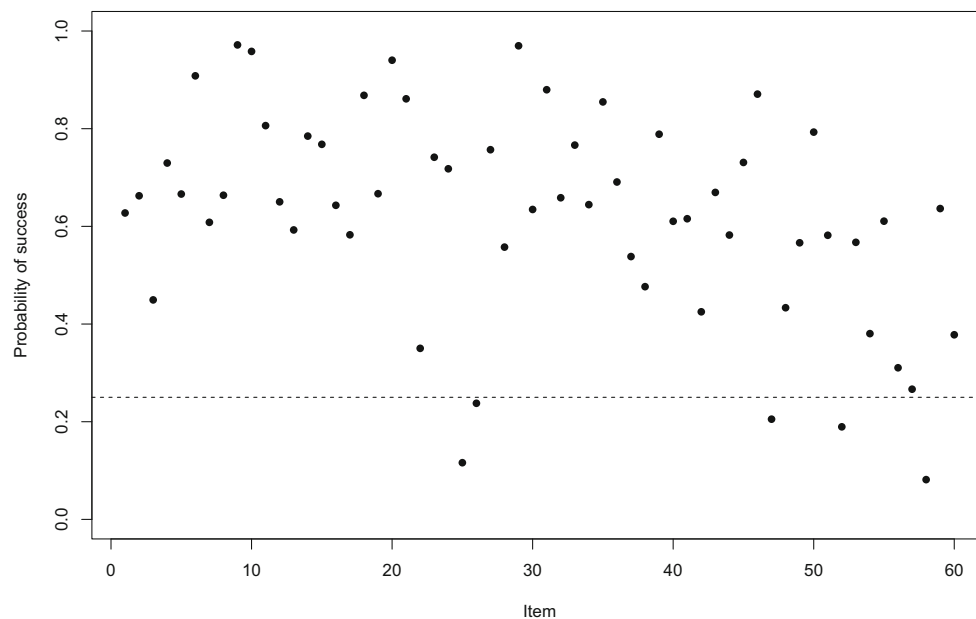
Item	$\alpha_i^{(R)}$		$\beta_i^{(R)}$		$\delta_i^{(S_1)}$		$\gamma_i^{(S_1)}$	
	EST	SE	EST	SE	EST	SE	EST	SE
31	0.45	0.16	0.34	0.11	-1.20	0.16	-0.43	0.35
32	0.05	0.12	0.17	0.09	-0.66	0.13	0.52	0.39
33	0.43	0.14	0.14	0.10	-0.93	0.13	-0.43	0.37
34	0.45	0.16	0.19	0.11	-1.99	0.22	0.88	0.32
35	0.46	0.16	-0.01	0.10	-0.69	0.13	0.30	0.37
36	0.20	0.13	0.02	0.10	-0.68	0.14	-1.06	0.38
37	0.44	0.15	0.00	0.10	-1.29	0.18	-0.10	0.38
38	0.42	0.14	-0.07	0.09	-0.62	0.13	-0.28	0.30
39	0.07	0.16	0.04	0.11	-0.99	0.15	-0.70	0.40
40	0.81	0.19	-0.20	0.11	-1.19	0.15	-0.25	0.45
41	0.59	0.16	-0.17	0.10	-1.05	0.14	-1.10	0.47
42	0.46	0.15	-0.16	0.09	-0.71	0.13	-0.05	0.43
43	0.45	0.14	-0.34	0.10	-1.34	0.15	-0.55	0.35
44	0.35	0.16	-0.37	0.10	-1.89	0.24	-0.96	0.40
45	0.73	0.18	-0.61	0.12	-1.77	0.19	-0.87	0.35
46	0.32	0.14	-0.36	0.10	-1.00	0.16	0.39	0.38
47	0.58	0.18	-0.60	0.11	-1.43	0.20	-0.07	0.53
48	0.57	0.17	-0.81	0.12	-1.91	0.20	-0.40	0.49
49	0.72	0.18	-1.52	0.15	-3.53	0.37	-2.02	0.41
50	0.51	0.18	-1.60	0.14	-2.75	0.27	-0.85	0.38
51	0.21	0.15	-0.71	0.10	-1.32	0.15	-0.69	0.38
52	0.39	0.14	-0.64	0.10	-0.80	0.13	0.49	0.33
53	0.04	0.13	0.31	0.09	0.62	0.13	1.33	0.37
54	1.11	0.33	-3.17	0.34	-3.47	0.34	-0.55	0.41
55	0.77	0.28	-2.62	0.24	-3.13	0.33	-0.70	0.43
56	0.48	0.17	-1.50	0.13	-1.82	0.18	0.45	0.48
57	-0.25	0.15	1.05	0.10	1.35	0.15	-0.03	0.38
58	0.20	0.15	-0.59	0.10	-0.59	0.13	-0.15	0.37
59	-0.23	0.19	1.71	0.13	2.42	0.22	0.28	0.34
60	0.45	0.25	-2.51	0.19	-2.29	0.20	0.26	0.39

$\alpha_i^{(R)}$  and  $\beta_i^{(R)}$  are the item discrimination and minus intercept parameter estimates for the ability class and  $\delta_i^{(S_1)}$  are minus the intercept parameter estimates from the secondary (non-ability) class.  $\gamma_i^{(S_1)}$  are the regression coefficients of response time on the probability of being classified into class  $S_1$ . (Continued)

the  $\gamma_i^{(S_g)}$  and  $\delta_i^{(S_g)}$  parameters between the two secondary classes ( $g = 1, 2$ ). That is, we imposed  $\delta_i^{(S_1)} = \delta_i^{(S_2)} + c_1$  and  $\gamma_i^{(S_1)} = \gamma_i^{(S_2)} + c_2$ , where superscripts ( $S_1$ ) and ( $S_2$ ) represent the first and second secondary classes (class 2 and class 3), respectively and  $c_1$  and  $c_2$  represent constants. The constraints that we imposed in the three class model were devised based on the following motivations: First, imposing these constraints is a confirmatory approach to investigate the two secondary classes as being different with respect to the level of performance (which is a possibility that we have considered earlier in this paper). Second, the number of model parameters is largely reduced in such a way

to improve interpretations. Accordingly, we hypothesized that there might be systematic differences in the model parameters ( $\gamma_i^{(S_g)}$  and  $\delta_i^{(S_g)}$ ) between the two secondary classes. Researchers, however, may consider other types of constraints based on their theoretical and/or practical needs.

The difference in the  $\delta_i^{(S_g)}$  parameters ( $c_1$ ) was estimated as 0.536 (SE = 0.054), while the difference in the  $\gamma_i^{(S_g)}$  parameters ( $c_2$ ) was nearly zero. Hence, we further constrained the  $c_2$  constant (between-class difference in the  $\gamma_i^{(S_g)}$  parameters) to be zero. The fit of this constrained three-class RT-based model (Model B3\*) was improved compared to the two-class RT-based model (Model B2) (see



**Fig. 4** Probability of success for the 60 test items for those respondents who were identified in the secondary class ( $S_1$ ) from the two-class RT-based model (Model B2)

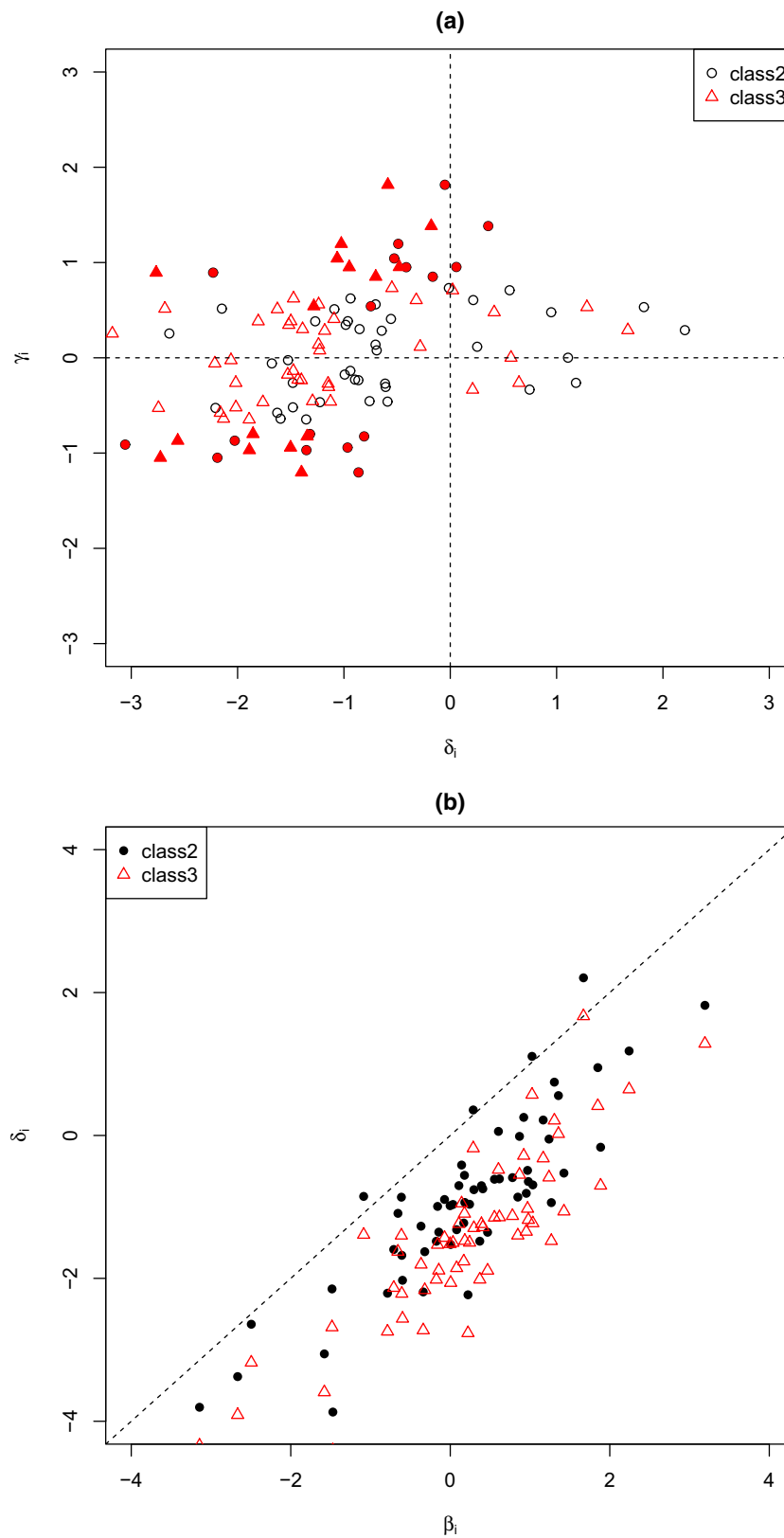
Table 1 for the fit statistics). Incorporating an additional secondary class (i.e., four-class model; Model B4) did not meaningfully improve the goodness-of-fit of the constrained three-class model in terms of the SABIC measure (see Table 1).

We proceeded to examine the characteristics of the constrained three-class model (Model B3\*). Figure 5 displays: (a) a comparison between the  $\delta_i^{(S_g)}$  difficulty parameter estimates and the  $\gamma_i^{(S_g)}$  parameter estimates for the two secondary classes and (b) a comparison between the  $\beta_i^{(R)}$  parameter estimates for the regular class and the  $\delta_i^{(S_g)}$  difficulty parameter estimates for the two secondary classes all parameter estimates and their standard errors are provided in Tables 4 and 5). When interpreting the figures, note that although Model B3\* includes two secondary classes, the item difficulty parameter ( $\delta_i^{(S_g)}$ ,  $g = 1, 2$ ) and the regression coefficients ( $\gamma_i^{(S_g)}$ ,  $g = 1, 2$ ) are equivalent for the two secondary classes, while for  $\delta_i^{(S_g)}$  there is a constant difference between the two secondary classes parameters (as explained above).

The patterns from Fig. 5 were generally similar to the patterns observed in Fig. 3. The  $\gamma_i^{(S_g)}$  coefficients tended to be positively associated with the  $\delta_i^{(S_g)}$  difficulty estimates (with a correlation of 0.49 for all items and 0.72 for the significant  $\gamma_i^{(S_g)}$  items). The correlations are not surprising because class 3 is class 2 plus an additive constant and classes 3 and 2 are basically class 2 from the previous, two-class RT-based model (Model B2), as will be explained with

percentages in the next paragraph. From Fig. 5b, we also observe that the  $\delta_i^{(S_g)}$  difficulty parameter estimates for the two secondary classes tended to be smaller than the  $\beta_i^{(R)}$  parameter estimates for the regular class. That is, the items appeared to be generally easier for the respondents who did not use their ability compared to those respondents who indeed used their abilities in item solving. The two secondary classes were differentiated in terms of their overall success rates. On average, the class 3 examinees were more successful than the class 2 examinees in the probability of success, while the difference was significantly different at the 5% level (0.536, SE= 0.054).

The constrained three-class model (Model B3\*) showed clear differentiation of latent classes; 61.6% were identified in the ability class, 15.4% in the first secondary class and 23% in the second secondary class. It is noteworthy that with the application of the three-class model, most examinees who were initially identified in the secondary class from the two-class model were re-classified into the two secondary classes. Specifically, 41.4% (154 out of 372) were classified into the first non-ability class and 56.5% (210 out of 372) into the second non-ability class (the remaining 2.1% was assigned to the ability class.), where the second non-ability class showed a higher mean compared to the first non-ability class. This result indicates that the no-ability class from the two-class model was indeed decomposed into two groups with the three-class model. This result supports our hypothesis that multiple non-ability-based classes existed that differed in their overall success rates.



**Fig. 5** (a) Scatter plot of  $\delta_i^{(S_g)}$  and  $\gamma_i^{(S_g)}$  (filled red dots and triangles for the items with significant  $\gamma_i^{(S_g)}$  estimates) and (b) scatter plot of the  $\beta_i^{(R)}$  parameter estimates from the regular class and the  $\delta_i^{(S_g)}$  parameter estimates from the two secondary classes (class 2 ( $S_1$ ) and class 3 ( $S_2$ )) from the three-class RT-based model (Model B3\*)

**Table 4** Item parameter estimates (for Items 1 to 30) from the constrained three-class RT-based model (Model B3\*, with an ability class and two secondary classes)

Item	$\alpha_i^{(R)}$		$\beta_i^{(R)}$		$\delta_i^{(S_1)}$		$\gamma_i^{(S_1)}$	
	EST	SE	EST	SE	EST	SE	EST	SE
1	0.36	0.16	1.30	0.11	0.75	0.14	-0.33	0.35
2	0.33	0.17	-1.09	0.11	-0.85	0.13	0.30	0.63
3	0.29	0.28	2.24	0.17	1.18	0.16	-0.26	0.34
4	-0.08	0.15	0.60	0.09	0.06	0.14	0.95	0.40
5	-0.16	0.24	1.85	0.14	0.95	0.15	0.48	0.39
6	0.73	0.33	3.19	0.26	1.82	0.19	0.53	0.32
7	-0.20	0.15	1.36	0.11	0.56	0.15	0.71	0.41
8	0.13	0.16	0.92	0.10	0.25	0.13	0.12	0.39
9	0.07	0.14	1.16	0.11	0.22	0.14	0.61	0.49
10	0.26	0.16	0.87	0.10	-0.01	0.13	0.73	0.39
11	0.12	0.17	1.24	0.11	-0.05	0.12	1.82	0.48
12	0.03	0.21	1.88	0.14	-0.16	0.15	0.85	0.36
13	0.20	0.18	1.42	0.12	-0.53	0.15	1.04	0.42
14	0.15	0.15	0.97	0.10	-0.49	0.13	1.20	0.42
15	0.08	0.16	0.98	0.11	-0.64	0.15	0.28	0.41
16	0.36	0.16	1.03	0.11	-0.69	0.14	0.08	0.41
17	0.14	0.16	0.95	0.10	-0.81	0.15	-0.82	0.41
18	0.26	0.17	0.62	0.10	-0.61	0.14	-0.31	0.35
19	0.47	0.14	0.78	0.10	-0.59	0.13	-0.46	0.40
20	0.27	0.18	1.27	0.12	-0.94	0.16	-0.14	0.42
21	0.36	0.14	0.55	0.10	-0.61	0.14	-0.27	0.41
22	0.58	0.17	0.85	0.12	-0.86	0.14	-1.20	0.42
23	0.18	0.13	0.14	0.09	-0.41	0.13	0.95	0.42
24	0.29	0.15	0.39	0.10	-0.70	0.13	0.14	0.48
25	0.39	0.14	0.41	0.10	-0.75	0.13	0.54	0.27
26	0.46	0.15	0.30	0.10	-0.76	0.14	-0.46	0.31
27	0.72	0.18	0.18	0.10	-0.56	0.13	0.41	0.45
28	0.23	0.14	0.47	0.10	-1.35	0.17	-0.65	0.42
29	0.15	0.13	0.24	0.09	-0.96	0.15	0.38	0.35
30	0.48	0.17	0.11	0.10	-0.70	0.13	0.56	0.47

$\alpha_i^{(R)}$  and  $\beta_i^{(R)}$  are the item discrimination and minus intercept parameter estimates for the ability class and  $\delta_i^{(S_1)}$  are minus the intercept parameter estimates from the first secondary (non-ability) class.  $\gamma_i^{(S_1)}$  are the regression coefficients of response time on the probability of being classified into class  $S_1$ . ( $\delta_i^{(S_2)}$  is obtained as  $\delta_i^{(S_2)} = \delta_i^{(S_1)} - 0.536$  and  $\gamma_i^{(S_2)} = \gamma_i^{(S_1)}$ .)

Figure 6 displays the probability of a correct response for the test takers who were identified in the two secondary classes.

Clearly, the response accuracy of the class 3 respondents was somewhat higher than the class 2 respondents. The probability of success for the class 3 and class 2 respondents was generally higher than .25, which is the accuracy level expected by chance for random guesses. This result confirms that the examinees identified in the two secondary classes may be described as successful knowledge retrievers rather than simple random guessers. The fact that there

were two such classes indicates that there were individual differences in the performance of the item-solving strategy that could be interpreted as knowledge retrieval.

## Simulation study

We conducted a simulation study to evaluate the accuracy of the estimated results described in section “Application”. Hence, we considered a two-class model (Model B2) and a constrained three-class model (Model B3\*) for a

**Table 5** Item parameter estimates (for Items 31 to 60) from the constrained three-class RT-based model (Model B3\*, with an ability class and two secondary classes)

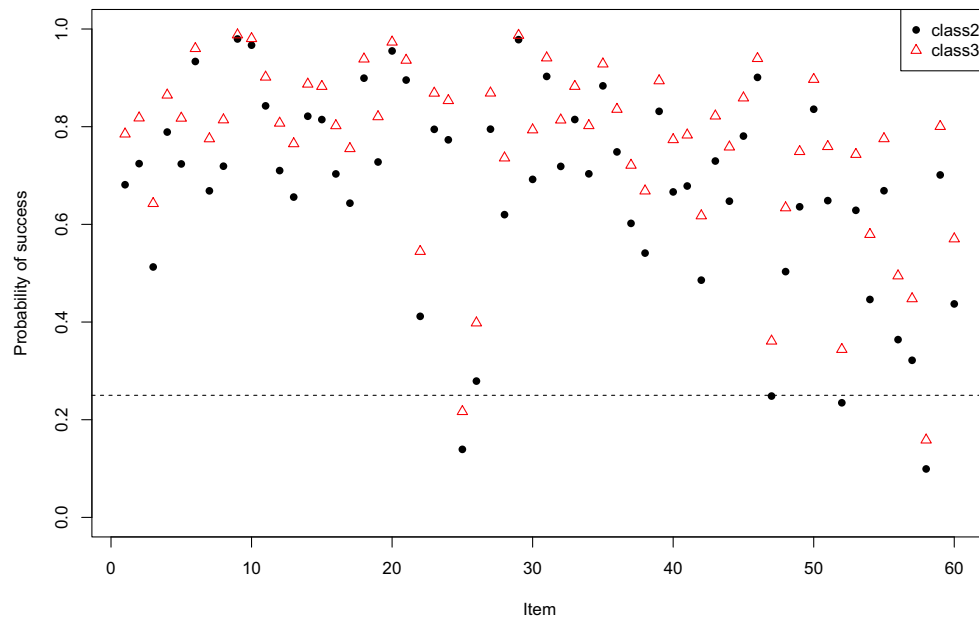
Item	$\alpha_i^{(R)}$		$\beta_i^{(R)}$		$\delta_i^{(S_1)}$		$\gamma_i^{(S_1)}$	
	EST	SE	EST	SE	EST	SE	EST	SE
31	0.44	0.16	0.37	0.10	-1.48	0.17	-0.52	0.33
32	0.05	0.12	0.18	0.09	-0.94	0.14	0.62	0.38
33	0.42	0.15	0.17	0.10	-1.23	0.14	-0.47	0.40
34	0.44	0.16	0.22	0.11	-2.23	0.22	0.89	0.30
35	0.47	0.16	-0.00	0.10	-0.98	0.13	0.35	0.39
36	0.18	0.13	0.03	0.09	-0.97	0.14	-0.94	0.39
37	0.48	0.15	0.00	0.10	-1.53	0.17	-0.02	0.40
38	0.44	0.15	-0.07	0.10	-0.90	0.14	-0.23	0.33
39	0.03	0.16	0.08	0.10	-1.32	0.17	-0.80	0.38
40	0.81	0.19	-0.17	0.11	-1.48	0.15	-0.26	0.47
41	0.59	0.16	-0.14	0.10	-1.35	0.15	-0.97	0.47
42	0.45	0.16	-0.16	0.09	-0.99	0.14	-0.18	0.46
43	0.46	0.15	-0.32	0.10	-1.63	0.16	-0.58	0.35
44	0.32	0.16	-0.34	0.10	-2.19	0.25	-1.05	0.43
45	0.74	0.20	-0.60	0.13	-2.03	0.19	-0.87	0.38
46	0.34	0.13	-0.37	0.10	-1.27	0.15	0.38	0.35
47	0.61	0.17	-0.61	0.11	-1.68	0.18	-0.06	0.45
48	0.56	0.17	-0.79	0.12	-2.21	0.19	-0.53	0.42
49	0.69	0.18	-1.47	0.14	-3.87	0.38	-1.97	0.41
50	0.50	0.19	-1.58	0.14	-3.06	0.27	-0.91	0.35
51	0.22	0.16	-0.71	0.10	-1.60	0.16	-0.64	0.36
52	0.42	0.15	-0.66	0.10	-1.09	0.13	0.51	0.30
53	0.08	0.13	0.29	0.09	0.36	0.13	1.38	0.36
54	1.11	0.32	-3.14	0.33	-3.80	0.33	-0.56	0.39
55	0.84	0.28	-2.67	0.25	-3.38	0.30	-0.64	0.45
56	0.47	0.17	-1.48	0.13	-2.15	0.18	0.52	0.43
57	-0.24	0.16	1.02	0.10	1.11	0.15	0.00	0.36
58	0.25	0.15	-0.61	0.10	-0.86	0.13	-0.23	0.39
59	-0.19	0.18	1.67	0.13	2.21	0.22	0.29	0.32
60	0.44	0.25	-2.50	0.19	-2.64	0.20	0.26	0.40

$\alpha_i^{(R)}$  and  $\beta_i^{(R)}$  are the item discrimination and minus intercept parameter estimates for the ability class and  $\delta_i^{(S_1)}$  are minus the intercept parameter estimates from the first secondary (non-ability) class.  $\gamma_i^{(S_1)}$  are the regression coefficients of response time on the probability of being classified into class  $S_1$ . ( $\delta_i^{(S_2)}$  is obtained as  $\delta_i^{(S_2)} = \delta_i^{(S_1)} - 0.536$  and  $\gamma_i^{(S_2)} = \gamma_i^{(S_1)}$ .) (Continued)

testing situation analogous to the empirical study setting. Specifically, 60 test items were considered for 1000 subjects, where response time was generated based on a log-normal model, such as  $\log(RT_{ip}) = \nu + \epsilon_{ip}$ , where  $\nu$  is the overall mean (set to -4, similar to the empirical study) and  $\epsilon_{ip} \sim N(0, 1)$ . The data-generating parameter values were set similar to the parameter estimates obtained from the two-class model fitted to the empirical data. We additionally considered a larger sample size condition with 5000 subjects to compare its parameter recovery results with the  $N = 1000$  case.

The mixing proportions of the two models were determined by the regression coefficients of the response time predictors of the secondary classes. For the two-class model (Model B2), the mixing proportions were  $\pi_1 = 0.62$  and  $\pi_2 = 0.38$ , while for the constrained three-class model (Model B3\*), they were  $\pi_1 = 0.62$ ,  $\pi_2 = 0.23$ , and  $\pi_3 = 0.15$ . For each condition, 100 datasets were generated and estimated with Mplus. The same maximum likelihood estimation setting was used as in the empirical study.

Note that label switching is of minimal concern in our simulation studies; for the two-class model there is



**Fig. 6** Probability of success for the 60 test items for those respondents who were identified in the two secondary classes (class 2 ( $S_1$ ) and class 3 ( $S_2$ )) from the three-class RT-based model (Model B3\*)

asymmetry between the two response strategy classes; that is, the two classes have different model structures in that one class includes an ability component (a regular ability based strategy class) and the other class does not (knowledge retrieval class). There is also asymmetry between two secondary classes for the constrained three class model because of the presence of the constant parameter  $c_1$  in one secondary class.

### Results: Two-class model

Figures 7 and 8 display the bias and root mean square error (RMSE) of the estimated item parameters from the two-class model (Model B2) under two sample size conditions.

Overall, the bias and RMSE of the estimated parameters were minor for most parameters and in all cases, the 95% confidence intervals of the bias included 0. For the  $\gamma^{(S_1)}$  parameters (regression coefficients of the response time predictors), the bias and RMSE were somewhat large compared to other parameters especially when the sample size is  $N = 1000$ . However, the bias of  $\gamma^{(S_1)}$  greatly decreased when  $N = 5000$ . The RMSE reduction was outstanding as shown in Fig. 8.

We also assessed classification accuracy with two measures: (1) the proportion of simulated datasets with correct classification per person, and (2) Cohen's kappa coefficient (which is a measure of agreement between categorical variables). Based on these two statistics, we

found the classification accuracy generally excellent for the two-class model in all sample sizes. When  $N = 1000$ , the proportion of correct classification ranged from 82% to 99% (with a mean of 92.6%) and Cohen's kappa coefficient ranged from 0.78 to 0.88 (with a mean of 0.84) across the respondents. When  $N = 5000$ , the proportion of correct classification ranged from 82 to 100% (with a mean of 93.3%) and Cohen's kappa coefficient ranged from 0.83 to 0.88 (with a mean of 0.86) across the respondents.

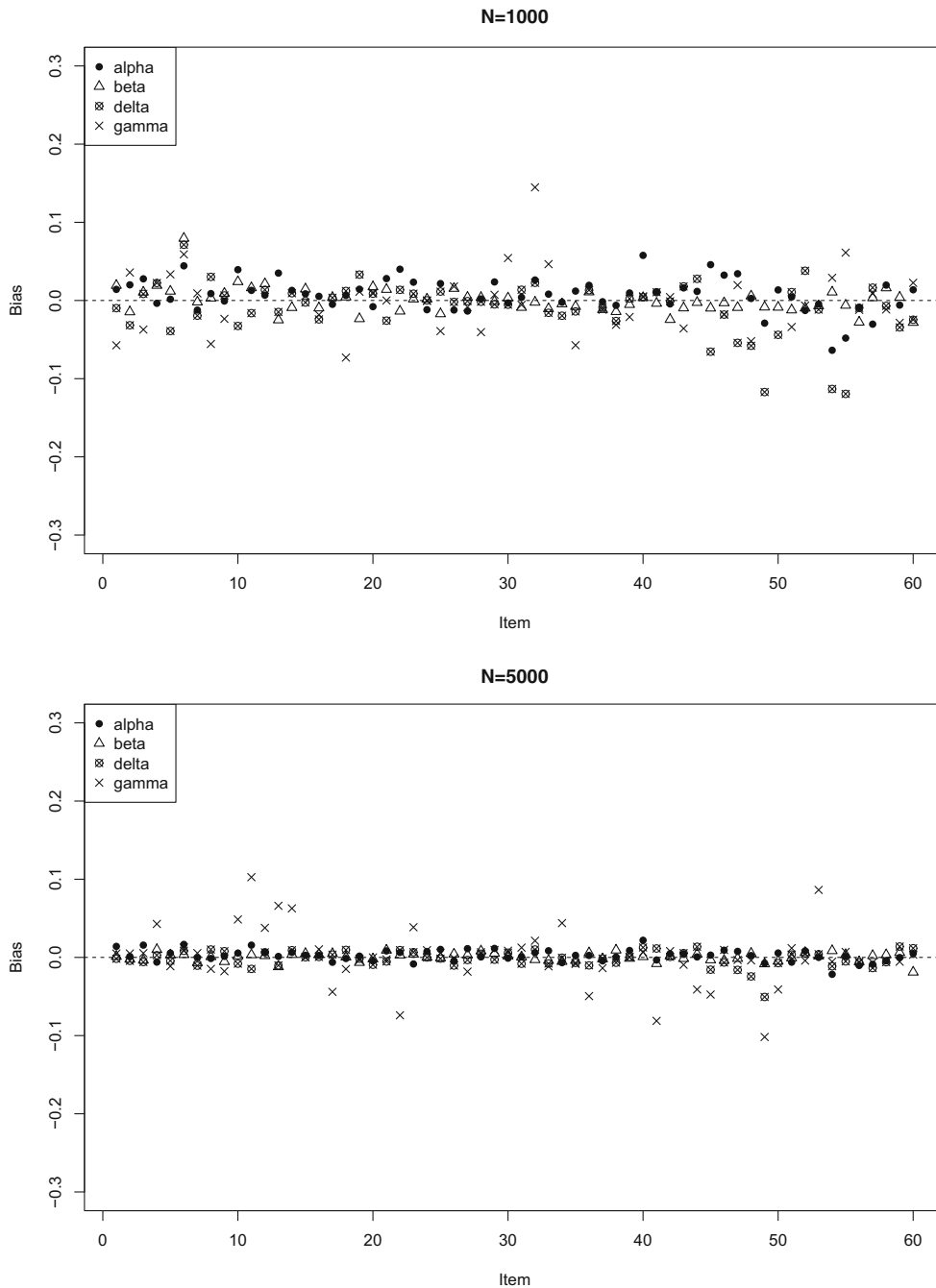
### Results: Three-class model

Figures 9 and 10 display the bias and RMSE of the estimated item parameters from the constrained three-class model (Model B3\*) in the three sample size conditions.

Similar to the two-class model case, the bias and RMSE of the estimated parameters were generally minor for most parameters except for the  $\gamma^{(S_g)}$  parameters when  $N = 1000$ . However, the bias and RMSE greatly decreased when the sample size was  $N = 5000$ . In addition, the 95% confidence intervals of the bias included 0 in all cases when  $N = 1000$  and  $N = 5000$ .

Classification accuracy was great for the constrained three class model in both sample sizes. When  $N = 1000$ , the proportion of correct classification ranged from 70 to 100% (with a mean of 87.0%) and Cohen's kappa coefficient ranged from 0.76 to 0.83 (with a mean of 0.79) across the 1000 subjects. When  $N = 5000$ , the proportion of correct





**Fig. 7** Bias of the estimated  $\alpha_i^{(R)}$ ,  $\beta_i^{(R)}$ ,  $\delta_i^{(S_1)}$ , and  $\gamma_i^{(S_1)}$  parameter values for the two-class RT-based model (Model B2) when  $N = 1000$  and  $N = 5000$

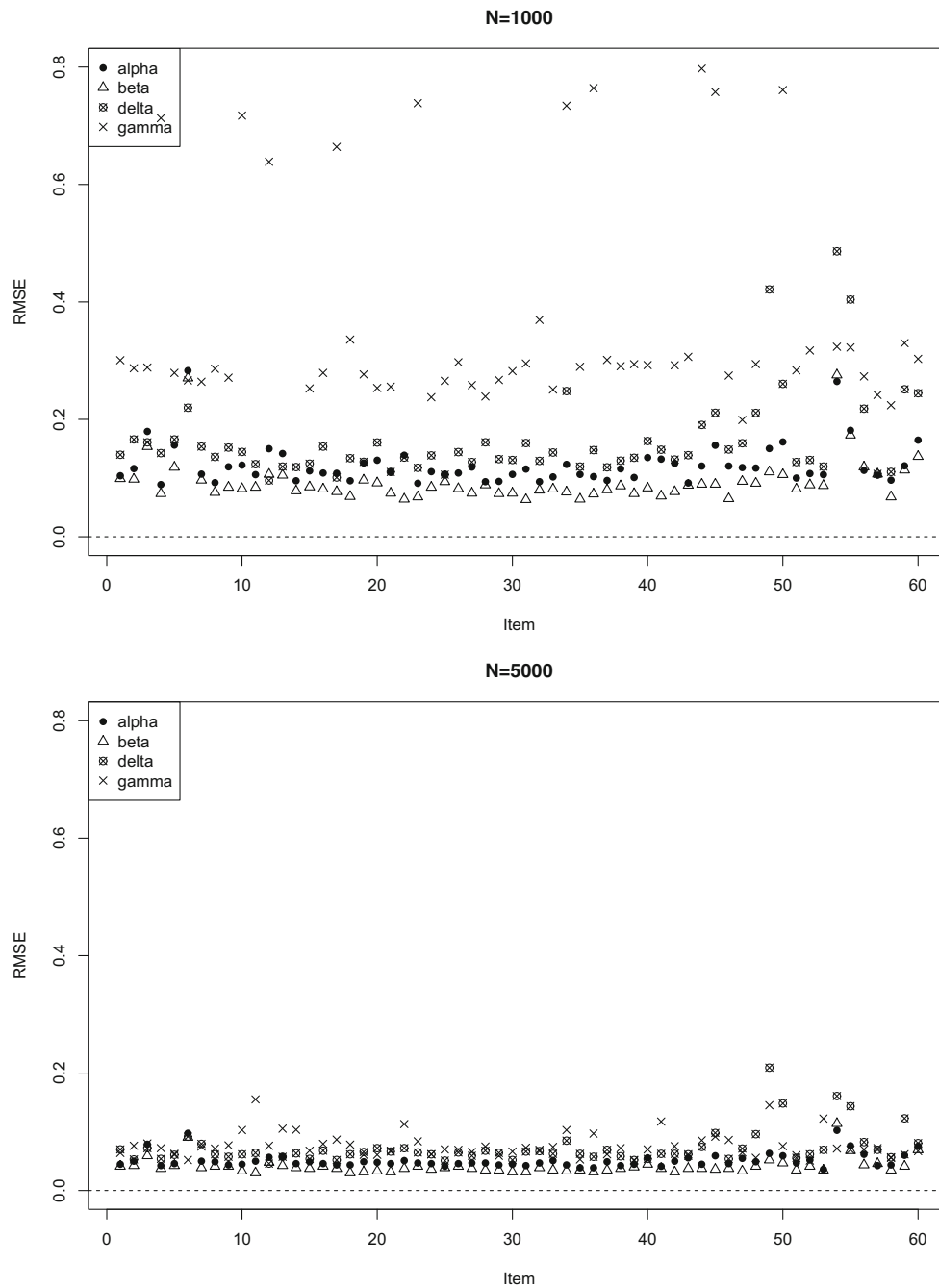
classification ranged from 70 to 100% (with a mean of 87.7%) and Cohen’s kappa coefficient ranged from 0.80 to 0.82 (with a mean of 0.81) across the 5000 subjects.

### Discussion

The present paper is a follow-up study to earlier investigations on respondents’ multiple item response strategies. In

particular, in our study we have hypothesized and explored the kind of latent classes that result when response time per item was used as covariates of the mixture probabilities. Note that response time was centered in two ways such that the effects could not be interpreted as stemming from general speed differences between subjects and items.<sup>7</sup> Similar

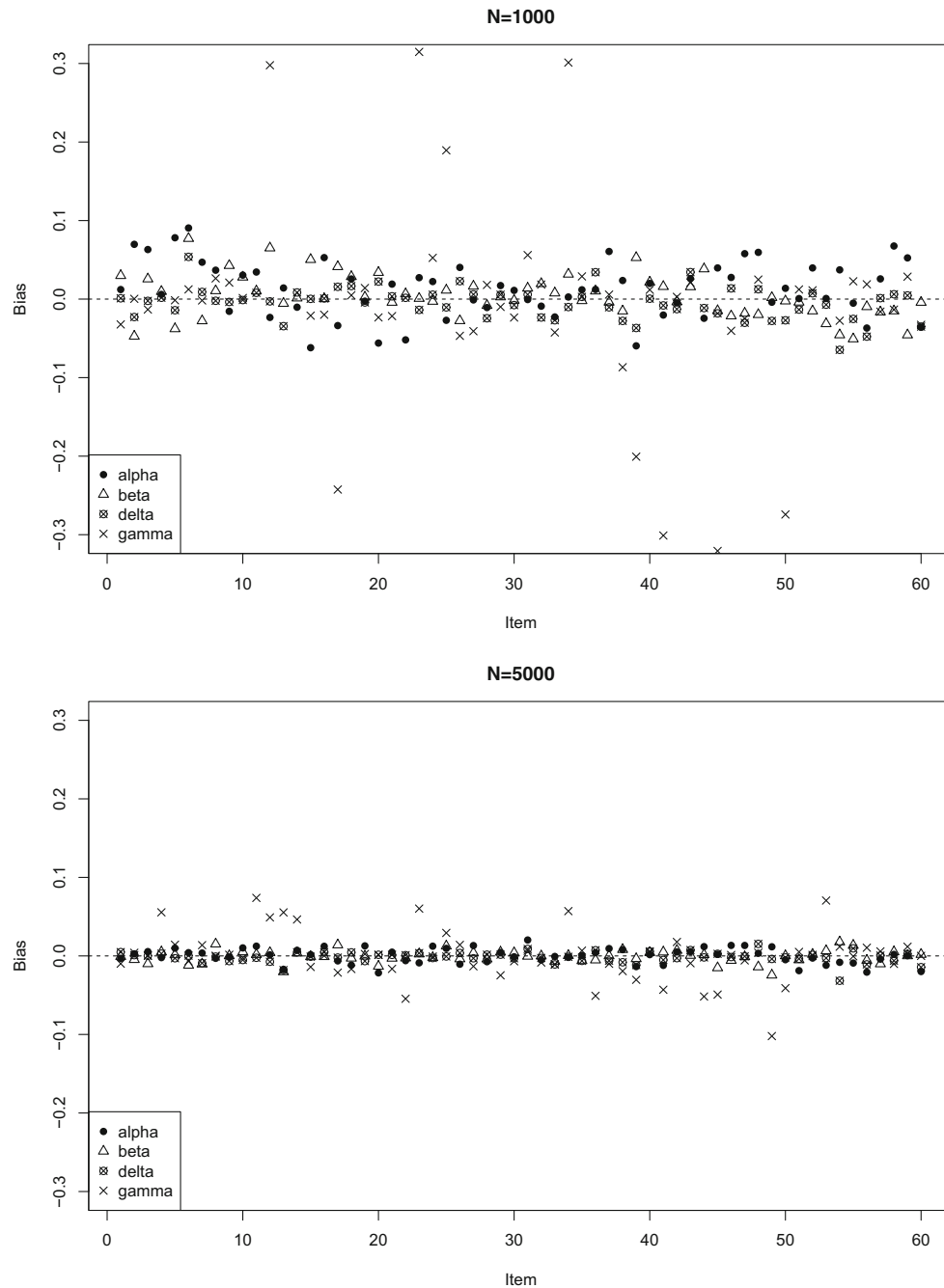
<sup>7</sup>We found that the results were similar without double centering the response time.



**Fig. 8** RMSE of the estimated  $\alpha_i^{(R)}$ ,  $\beta_i^{(R)}$ ,  $\delta_i^{(S_i)}$ , and  $\gamma_i^{(S_i)}$  parameter values for the two-class RT-based model (Model B2) when  $N = 1000$  and  $N = 5000$

to the existing multiple-strategy models, in our approach one of the two (or more) classes are classes without individual differences; thus, they are homogeneous classes without an ability component. For example, a guessing latent class would be a class with rather low probabilities of success. An alternative possibility as illustrated in our study is that the homogeneous class represents a simple knowledge retrieval

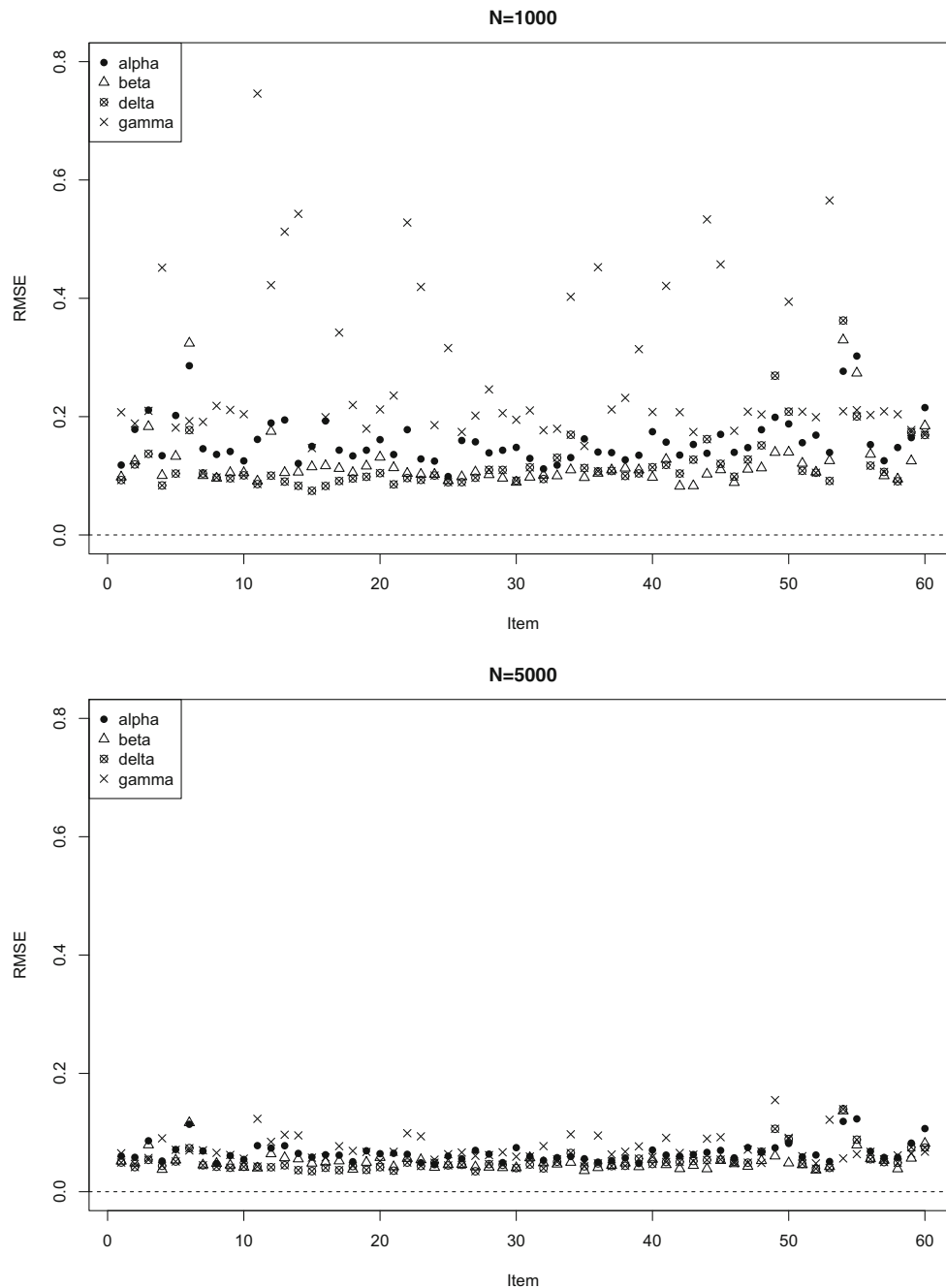
strategy instead of a regular, ability-based strategy. If knowledge retrieval is successful, the response would likely be given quickly. If knowledge retrieval does not work, one may expect a longer response time and a lower success rate. As a consequence, the alternative response strategy class would be characterized by fast responses for successful items as well as slower responses for less successful



**Fig. 9** Bias of the estimated  $\alpha_i^{(R)}$ ,  $\beta_i^{(R)}$ ,  $\delta_i^{(S_1)}$ , and  $\gamma_i^{(S_1)}$  parameter values for the constrained three-class RT-based model (Model B3\*) when  $N = 1000$  and  $N = 5000$ . The bias of the  $c_1$  parameter was  $-0.001$  and  $-0.0004$  when  $N = 1000$  and  $N = 5000$ , respectively

items. A third class can be included in the model if there would be differences among the knowledge retrieval strategy subjects in their overall knowledge levels and qualities. To our knowledge, a model with mixture probabilities based on item-specific effects of response times has not been used earlier to explore secondary latent classes, and the results are of a kind that a knowledge retrieval strategy is a possible explanation for secondary classes.

The analysis of our motivating data example, obtained from a low-stakes cognitive assessment without time pressure, showed that a non-trivial number of examinees employed possibly an item-response strategy based on knowledge retrieval rather than a random guessing strategy. We also found that two classes of potential knowledge retrieval could be differentiated based on their overall success level. Specifically, the hypothetical knowledge



**Fig. 10** RMSE of the estimated  $\alpha_i^{(R)}$ ,  $\beta_i^{(R)}$ ,  $\delta_i^{(S_1)}$ , and  $\gamma_i^{(S_1)}$  parameter values for the constrained three-class RT-based model (Model B3\*) when  $N = 1000$  and  $N = 5000$ . The RMSE of the  $c_1$  parameter was 0.0008 and 0.0002 when  $N = 1000$  and  $N = 5000$ , respectively

retrieval classes could be characterized as subjects with faster responses for easy items and slower responses for more difficult items. This type of strategy was likely to be different from the strategy based on item pre-knowledge, stemming from item compromise (e.g., Shu et al., 2013), given that the analyzed data came from a low-stakes assessment.

These results seem to show the presence of knowledge retrieval strategies that we hypothesized although additional work would be needed to further strengthen our hypothesis. We note that Bolsinova, De Boeck and Tijmstra (2017) analyzed the same dataset with a local dependence extension of the hierarchical model of van der Linden (2007) with a dimension for response accuracy and another

dimension for response time, where the local dependence refers to response time and correctness of the responses. Bolsinova et al. (2017) showed that after controlling for latent variables and item parameters, (1) response time is negatively related to the correctness of the response for most items, and (2) the dependence is related to item difficulty in that the dependence is positive for the most difficult items. This relationship between dependence and item difficulty is in line with our results. It is possible that the dependence stems from a subset of the subjects using a response strategy that consists of retrieving knowledge from surface memory for easier items while using less successful strategies for more difficult items. It was not our purpose to find an explanation for earlier findings regarding the same data set and neither did we aim at finding the true model. However, it is intriguing and encouraging that the results from both analyses seem to be consistent. One way to follow up on these results would be to analyze the data with a model that has two dependence-based latent classes: one with and another without residual dependence between response time and response accuracy.

As discussed earlier, our model is a meaningful extension of earlier multiple-strategy mixture modeling approaches (e.g., Mislevy & Verhelst, 1990), with response time covariates for the marginal probabilities of strategy classes. We utilized the response time to shed light on the nature of the secondary strategy classes. Response time information has also been used in the literature, often in the context of studying rapid guessing behavior. For instance, Schnipke and Scrams (1997) applied a finite-mixture model to response time data. Meyer (2010) and Wang and Xu (2015) applied a joint mixture model to both response data and response time, simultaneously. Wise and Demars (2006) utilized response time to determine time thresholds and to define speeded items. Molenaar et al. (2016) used response time information to define a class of fast responses and a class of slow responses. In a non-mixture modeling context, other researchers utilized response time as a predictor to account for heterogeneity in the probability of correct responses (e.g., Goldhammer et al., 2014; Goldhammer, Naumann, & Greiff, 2015; Roskam, 1997).

Our proposed modeling approach can be further improved in several ways. For instance, it could be beneficial to incorporate other behavioral, psychological, and even neural information in the model in order to investigate the nature of response strategy classes. In addition, although our model allows for within-subject differences in how item responses are generated, the differences between items are the same within each class. A more flexible model would be a mixture model for responses so that the strategies can vary within subjects

in different ways depending on the person. It may be meaningful to relax this assumption to allow for item-specific strategy choices as also discussed in e.g., Erosheva (2005) and Pokropek (2016). However, as noted earlier, it would imply that constraints of a different kind need to be imposed.

## Appendix

Here we provide an example Mplus code for fitting our proposed model (Model B2) to the motivating data example.

```
!! Header of input file
TITLE: Two-class mixture model for
!knowledge retrieval based item response
!strategy

!! Data file specification
DATA: FILE = example.dat;

!! Define variables and specify number
!of latent classes
VARIABLE:
NAMES ARE y1-y60 t1-t60; ! 60 item
!responses and 60 item response time
CATEGORICAL = y1-y60; ! binary responses
MISSING = ALL(99);
! missing data coded as 99
CLASSES = c (2);
! define # of latent classes

!! Estimation settings
ANALYSIS:
TYPE = MIXTURE; ! estimate
!finite mixture model
ALGORITHM = INTEGRATION; ! use
!integration method
!(with 15 default quadrature points)
STARTS = 20 10 ; ! random start

!! Model specification
MODEL:
! Overall model
%OVERALL%
th by y1-y60*; ! item loading
[th@0]; ! factor mean fixed at 0
th@1; !factor variance fixed at 1
! Item response time as predictors
```

```

c#1 on t1 t2 t3 t4 t5 t6 t7 t8 t9 t10
(d1-d10);
c#1 on t11 t12 t13 t14 t15 t16 t17 t18
t19 t20 (d11-d20);
c#1 on t21 t22 t23 t24 t25 t26 t27 t28
t29 t30 (d21-d30);
c#1 on t31 t32 t33 t34 t35 t36 t37 t38
t39 t40 (d31-d40);
c#1 on t41 t42 t43 t44 t45 t46 t47 t48
t49 t50 (d41-d50);
c#1 on t51 t52 t53 t54 t55 t56 t57 t58
t59 t60 (d51-d60);

```

```
[c#1@0] ; ! mean of class 1 fixed at 0
```

```
! Model for class 2 (ability group)
```

```
c#2%
```

```
th by y1-y60*; ! item loading
[th@0]; ! factor mean fixed at 0
th@1; !factor variance fixed at 1
[y1$1-y60$1] (i1-i60); ! difficulty
```

```
! Model for class 1 (secondary,
```

```
!no-ability group)
```

```
%c#1%
```

```
th by y1-y60@0; ! item loading
[th@0]; ! factor mean fixed at 0
th@0; !factor variance fixed at 0
[y1$1-y60$1] (g1-g60); ! difficulty
```

```
! Save posterior probabilities of
```

```
!latent class membership
```

```
SAVEDATA: FILE IS model_prob.txt;
```

```
SAVE = cprob;
```

## References

- Bennet, D., & Bennet, A. (2008). The depth of knowledge: Surface, shallow or deep? *Vine*, *38*, 405–420.
- Bloom, B., Engelhart, M., Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals handbook I: cognitive domain*. New York: David McKay Company.
- Bolsinova, M., De Boeck, P., & Tijmstra, J. (2017). Modelling conditional dependence between response time and accuracy. *Psychometrika*.
- Bolt, D. M., Cohen, A. S., & Wollack, J. A. (2002). Item parameter estimation under conditions of test speededness: Application of a mixture Rasch model with ordinal constraints. *Journal of Educational Measurement*, *39*, 331–348.
- Cao, J., & Stokes, S. (2008). Bayesian IRT guessing models for partial guessing behaviors. *Psychometrika*, *73*, 209–230.
- Chang, Y. W., Tsai, R. C., & Hsu, N. J. (2014). A speeded item response model: Leave the harder till later. *Psychometrika*, *79*, 255–274.
- Enders, C., & Tofghi, D. (2008). The impact of misspecifying class-specific residual variances in growth mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*, 75–95.
- Entwistle, N., & Peterson, E. (2004). Conceptions of learning and knowledge in higher education: Relationships with study behavior and inferences of learning environments. *International Journal of Educational Research*, *41*, 407–428.
- Erosheva, E. (2005). Comparing latent structures of the Grade of Membership, Rasch, and latent class models. *Psychometrika*, *70*, 619–628.
- Goldhammer, F., Naumann, J., Stelter, A., Toth, K., rolke, H., & klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology*, *106*, 608–626.
- Goldhammer, F., Naumann, J., & Greiff, S. (2015). More is not always better: The relation between item response and item response time in Raven's matrices. *Journal of Intelligence*, *3*, 21–40.
- Jong, T., & De Ferguson-Hessler, M. (1996). Types and qualities of knowledge. *Educational Psychologist*, *31*, 105–113.
- Meyer, J. (2010). A mixture Rasch model with item response-time components. *Applied Psychological Measurement*, *34*, 521–538.
- Mislevy, R. J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika*, *55*, 195–215.
- Molenaar, D., Oberski, D., Vermunt, J., & De Boeck, P. (2016). Hidden Markov IRT models for responses and response times. *Multivariate Behavioral Research*, *51*, 606–626.
- Pokropek, A. (2016). Grade of membership response time model for detecting guessing behaviors. *Journal of Educational and Behavioral Statistics*, *41*, 300–325.
- Roskam, E. E. (1997). Models for speed and time-limit tests. In van der Linden, W. J., & Hambleton, R. (Eds.) *Handbook of modern item response theory*, (pp. 87–208). New York: Springer.
- Schnipke, D. L., & Scrams, D. J. (1997). Modeling item response times with a two-state mixture model: A new method of measuring speededness. *Journal of Educational Measurement*, *34*, 213–232.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*, 127–190.
- Shu, Z., Henson, R., & Luecht, R. (2013). Using deterministic, gated item response theory model to detect test cheating due to item compromise. *Psychometrika*, *78*, 481–497.
- Tofghi, D. C. K. (2007). Enders Identifying the correct number of classes in mixture models. In Hancock, G. R., & Samulelsen, K. M. (Eds.) *Advances in latent variable mixture models*, (pp. 317–341). Greenwich: Information Age.
- van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*, *72*, 287–308.
- van der Linden, W. J. (2009). Conceptual issues in response-time modeling. *Journal of Educational Measurement*, *46*, 247–272.
- Wang, C. G., & Xu, G. (2015). A mixture hierarchical model for response times and response accuracy. *British Journal of Mathematical and Statistical Psychology*, *68*, 456–477.
- Wise, S. L., & Demars, C. E. (2006). An application of item response time: The effort moderated IRT model. *Journal of Educational Measurement*, *43*, 19–38.
- Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education*, *18*, 163–183.

- Yamamoto, K. (1989). *A HYBRID model of IRT and latent class models. ETS Research Report SeriesA HYBRID model of IRT and latent class models ETS Research Report Series*. Princeton: Educational Testing Service.
- Yamamoto, K. H. (1997). Everson Modeling the effects of test length and test time on parameter estimation using the HYBRID modelModeling the effects of test length and test time on parameter estimation using the HYBRID model. In Rost, J., & Langeheine, R.. (Eds.) *Applications of latent trait and latent class models in the social sciences*, (pp. 89-98). New York: Waxmann Verlag GmbH.
- Yang, C. C. (2006). Evaluating latent class analysis models in qualitative phenotype identification. *Comput Stat Data Anal*, 50, 1090–1104.