Article

Cognitive Trait Model: Measurement Model for Mastery Level and Progression of Learning

Jaehwa Choi

Department of Educational Leadership, Graduate School of Education and Human Development, The George Washington University, Washington, DC 20010, USA; jaechoi@gwu.edu

Abstract: This paper seeks to establish a framework which operationalizes cognitive traits as a portion of the predefined mastery level, the highest level expected to successfully perform all of the relevant tasks of the target trait. This perspective allows us to use and interpret the cognitive trait levels in relative quantities (e.g., %s) of the mastery level instead of relative standings (i.e., rankings) on an unbounded continuum. To facilitate the proposed perspective, this paper presents an analytical framework that has support on the [0, 1] trait continuum with truncated logistic link functions. The framework provides a solution to cope with the chronic question of "relative standings or magnitudes of learning outcome?" in measuring cognitive traits. The proposed framework is articulated relative to the traditional models and is illustrated with both simulated and empirical datasets within the Bayesian framework, estimated with the Markov chain Monte Carlo method.

Keywords: cognitive trait model; latent variable model; factor model; item response theory model; logistic function; truncated logistic function; Bayesian inference; Markov chain Monte Carlo

MSC: 62P15

1. Introduction

The primary interest of the social and behavioral sciences is typically the *traits* of human beings, where the developments of tools to model these traits are of critical theoretical and practical interest. Messick defined a trait as "a relatively stable characteristic of a person . . . which is consistently manifested to some degree when relevant, despite considerable variation in the range of setting and circumstances" [1]. According to [2], the social and behavioral sciences encounter two major domains of human traits: *cognitive* and *affective*. In the cognitive domain, researchers would like to model a wide range of traits:

Various classes of cognitive and neuropsychological functioning, including intelligence, broad ability domains, and more focused domains (e.g., abstract reasoning and categorical thinking; academic achievement; attention; cognitive ability; executive function; language; learning and memory; motor and sensorimotor functions and lateral preferences; and perception and perceptual organization/integration) [3] (p. 155).

In the affective domain, on the other hand, the researchers are customarily interested in feelings or emotions, such as attitude or self-efficacy. Although the field of psychometrics covers both of the domains, in educational measurement or learning science specifically, the traits in the cognitive domain have been of primary interest.

Ref. [4], for example, defined the primary goal of educational and psychological measurement as "the determination of how much of such a latent ability (trait) a person possesses" (p. 3). They continued, "The underlying idea here is that, if one could physically ascertain the ability of a person, this ruler would be used to tell how much ability a given person has, the ability of several persons could be compared" (p. 3). As such, the current work will be framed largely within the cognitive domain.

A cognitive trait has been conceptualized as an *ability* or *proficiency* to perform relevant tasks to some degree (e.g., [4]). What does it mean to have the ability to do relevant tasks?

What does it mean to have the ability to recall the brand of the phone you are using, to compute 2 + 2, or to ride a bicycle? One natural answer could be "to have an ability is to *know how* to do some relevant tasks to some degree". To have the ability to name the brand of a phone is to know how to name the brand of a phone, to have the ability to compute 2 + 2 is to know how to solve 2 + 2, and to have the ability to ride a bicycle is to know how to ride a bicycle. In other words, a cognitive trait (i.e., the ability to do or knowledge how to do) is manifested in performing the relevant tasks identified, regardless of whether the target knowledge is related to knowledge-how or knowledge-that.

From the observations of the performance of related tasks, we infer the levels of the target cognitive trait, that is, "how much ability a learner has acquired (or possessed) to perform relevant tasks?" Therefore, to appropriately infer a cognitive trait level, the features of the trait and their interdependencies should be agreed among the relevant experts and carefully melted into these tasks. In other words, the domain (i.e., a complete list) of tasks must be pre-defined to measure (especially, to measure) a cognitive trait.

Given a domain of tasks for a cognitive trait, we can naturally assume there is a cognitive trait level that is expected to perform *all* of those tasks successfully. This, as the highest expected level, should be called the *mastery level*. Under the same logic, there is a level that is expected where one fails at all of those tasks, which is the lowest-level, and which should be called the *ignorance level*. Then, the levels between these two boundaries represent the sizes of the portions (parts) of the mastery. This induces an analytical framework that maps the trait levels on a *bounded* continuum from the lowest level to the highest (i.e., mastery) level. Consider an instrument of measurement (an examination or test) for a problem-solving trait: "For any number from 1 to 9, find the number that makes 10 when added to the given number" [5]. Let us further assume that there are only 10 tasks (i.e., the 10 tasks are the domain of the tasks) for the trait. Ignoring the measurement error and bias for the moment, if a person fails to solve all of the 10 tasks, we would infer that this person is at the lowest trait level. Likewise, if a person successfully solves all of the 10 tasks (again temporarily ignoring the issue of measurement error and bias), we would conclude that this person is at the mastery level.

To model the connections between the levels of cognitive (or affective) domain traits (constructs or latent variables) and the observations from relevant tasks (items or indicators), latent variable models (LVMs) are especially useful. LVMs include the factor models (FMs), structural equation models (SEMs), and item response theory models (IRTMs). In traditional LVMs, the trait levels have typically been assumed to be on the *unbounded* real line continuum (i.e., from negative infinity to positive infinity), with a midpoint of zero. This assumption, however, may be more due to the nature of the mathematical model being employed than a theoretical belief that the levels of a trait can keep decreasing or increasing indefinitely.

Although the traditional analytical framework has been proved to be useful in analyzing the relationship among constructs or to measure the relative standing levels of cognitive traits for the single point, group comparison, and/or summative assessment purposes, it has limitations in continuous (i.e., longitudinal) and/or individually adaptive assessments, such as formative assessment, the learning growth model, or real-time learning analytics in the digital environment. Suppose that someone's trait level is estimated as −1.5, with an unbounded real line continuum. Although the (unbounded) continuum value is useful for placing and differentiating the individuals' relative location (i.e., ranking), it does not directly provide information on the level of achievement or acquisition compared to a mastery level before making additional efforts (some additional transformation techniques will be discussed later in this paper).

In fact, researchers have proposed several options that better conceptualize and represent trait estimates differently. Perhaps the most popular, and simplistic, transforms the latent variable estimates to a distribution with a designated mean and standard deviation (e.g., $\mu = 100$ and $\sigma = 15$, as for some intelligence test scores). Within the IRTM arena, ref. [6] suggested a method using the number of correct scores as the basis for the item

characteristic curves (ICCs) rather than the trait score; however, this approach was not further developed, because the ICCs were dependent on the set of items included in a test (i.e., test-specific). Reference [7] suggested a domain score that is an expected number-correct score, derived by transforming an examinee's IRTM trait estimate via a weighted sum of each ICC. The authors argued that the domain scores are preferable in the context of student qualification. Although this domain score approach enhances the interpretation of the assessment result, it is still an *indirect* interpretation of the trait estimates and requires transformations with IRT estimation results. More complex approaches, within IRTMs, Item Maps, or Wright Maps [8] have been used to display both the person parameter (i.e., trait level) and the item difficulty parameter estimates along a common scale. Based on this approach, one can interpret a trait estimate of -1.5 as an examinee being expected to have a 0.5 probability of correctly answering an item whose difficulty is at -1.5 on the normal continuum (low difficulty). This interpretation, however, primarily describes the relation between the trait and items (again, the expected probability of correct answer) rather than the trait estimate itself, and it is still an indirect characterization/conceptualization of what the person parameter means.

Several LVMs to conceive of the latent variables as bounded have also been suggested, such as the Beta-Binomial model [9]. In this model, the person parameter represents a true score in the [0, 1] metric with a Beta prior distribution, and the model is used as a method for classifying the accuracy of the IRTMs [10]. Diagnostic classification models (DCMs; ref. [11]), which provide each subject with a profile detailing the attributes (latent variables), can also model latent categories. These models, however, are based on *discrete* traits—such as dichotomous (0 = no-mastery status or 1 = mastery status) or polytomous (0/1/2/ . . .) variables—to represent the stages of mastery and provide the expected probabilities of stage mastery. Another popular assessment modeling technique, Bayesian networks (BNs; ref. [12]), represent a set of latent variables and their conditional dependencies with a directed acyclic graph (DAG). However, in the same way as with CDMs, the BNs also model the latent variables as discrete, and provide *probability* estimates for each latent category.

For formative assessment, learning growth, or adaptive learning systems, which have rapidly gained in popularity in recent years, the researchers, practitioners, educators, and learners are in dire need of a new analytical framework that supports more direct and intuitive uses and interpretations of 'the magnitude of learning outcomes and its changes.' To help in the communication between the observations from the tasks and the trait continuum of interest in such uses and demands, therefore, it would be useful to have an analytical framework that facilitates the trait levels on the [0, 1] continuum as follows:

(1) Assigning zero (0) to the lower boundary (the lowest or ignorance level): a trait level as an expected proportion of the succeeded tasks out of all of the tasks is 0 (or 0%);
(2) Assigning unit (1) to the upper boundary (the highest or mastery level): a trait level for an expected proportion of the succeeded tasks out of all of the tasks is 1 (or 100%);
(3) Assigning a number between the 0 and 1 boundaries: a trait level as an expected proportion of the succeeded tasks out of all of the tasks.

This paper proposes an analytical framework for measuring cognitive traits (so-called, Cognitive Trait Model) based on truncated logistic functions to serve this purpose. First, analytical specifications of the traditional and proposed models are presented. Second, the various psychometric aspects of the proposed models are illustrated, alongside those of the traditional models. Third, a simulated dataset and an empirical dataset are analyzed to illustrate and evaluate the proposed model, using a Markov chain Monte Carlo (MCMC) estimation method [13]. Lastly, this paper concludes with a discussion and suggested future works.

2. Developments

2.1. Traditional Models

Let y_{ij} be the observed data of the ith person on the jth item (or task), and without loss of generality, let the possible values for each y_{ij} be 1 (indicating successful completion of

a task, a correct answer to an item, etc.) and 0 (indicating failure to complete a task, an incorrect answer to an item, etc.). IRTMs [14,15], widely used LVMs, can be expressed as a version of:

$$\Pr(y_{ij} = 1|\theta_i, a_j, b_j, c_j) = c_j + (1 - c_j)\Psi(a_j(\theta_i - b_j)), \tag{1}$$

where θ_i represents the level (score) of a cognitive trait for person i. The symbols a_j, b_j, and c_j represent the parameters for item j. Specifically, a_j is often referred to as the slope or *discrimination* parameter for the item, because it relates to how well an item differentiates (or discriminates) among the people with lower and higher values along the θ continuum. The symbol b_j represents the location parameter for the jth item, often referred to as the *difficulty* parameter for item j. Finally, c_j is often referred to as the *pseudo-guessing* parameter, reflecting the probability that even the least capable person will occasionally give a correct response, that is, $\Pr(y_{ij} = 1|\theta) > 0$, due to guessing (e.g., the probability of giving a correct response by probability is 0.25 on a selected-response item with four possible choices). The symbol Ψ, referred to as *link function* (or *item response function* in the context of IRTMs), is a monotonically increasing function that maps θ to the probability of $y_{ij} = 1$.

In traditional LVMs, the cumulative normal distribution function (i.e., normal ogive function) and the logistic function are popular choices for Ψ, with the latter being more common for IRTMs. In the case of choosing Ψ in Equation (1) to be a logistic function, this yields the familiar three-parameter logistic (3PL) IRTM [15]:

$$\Pr(y_{ij} = 1|\theta_i, a_j, b_j, c_j) = c_j + (1 - c_j) \times \frac{\exp(a_j(\theta_i - b_j))}{1 + \exp(a_j(\theta_i - b_j))}. \tag{2}$$

Setting $c_j = 0$, effectively assuming that guessing on items cannot yield a correct answer, yields the two-parameter logistic (2PL) IRTM [14]:

$$\Pr(y_{ij} = 1|\theta_i, a_j, b_j) = \frac{\exp(a_j(\theta_i - b_j))}{1 + \exp(a_j(\theta_i - b_j))}. \tag{3}$$

Finally, assuming all of the items have the same ability to discriminate among individuals, setting $a_j = 1$ yields the one-parameter logistic (1PL) IRTM or Rasch model [16]:

$$\Pr(y_{ij} = 1|\theta_i, b_j) = \frac{\exp(\theta_i - b_j)}{1 + \exp(\theta_i - b_j)}. \tag{4}$$

Whether logistic function or cumulative normal distribution function (i.e., normal ogive function), all of the models in Equations (2)–(4) are supported on an entire real number line for θ. In other words, the *support* (the set of θ points where the function is not zero-valued) of these functions is from $-\infty$ to $+\infty$. (for information on defining the latent distribution by assuming some non- or semi-parametric form, and simultaneously estimating the distribution with item parameters, see [17].

2.2. Proposed Models

In order to derive link functions supported on the [0, 1] continuum, that is, the levels of a cognitive trait assumed to be on a bounded 0 to 1 continuum ($0 \leq \theta \leq 1$), one can analytically transform the traditional 2-parameter logistic (2PL) IRTM function as follows. First, let z be the function from Equation (3):

$$z = \frac{\exp(a_j(\theta_i - b_j))}{1 + \exp(a_j(\theta_i - b_j))} \tag{5}$$

Subtracting 0.5 from z shifts the function to the left,

$$\frac{\exp(a_j(\theta_i - b_j))}{1 + \exp(a_j(\theta_i - b_j))} - \frac{1}{2}. \tag{6}$$

Creating a function that *passes* through the point $(b_j, 0)$ with the upper horizontal asymptote 0.5 and lower horizontal asymptote -0.5. Next, multiplying the shifted function by a factor δ stretches/shrinks the function in Equation (6) vertically:

$$\delta\left(\frac{\exp(a_j(\theta_i - b_j))}{1 + \exp(a_j(\theta_i - b_j))} - \frac{1}{2}\right). \tag{7}$$

Lastly, adding a constant ε shifts the above function vertically:

$$\delta\left(\frac{\exp(a_j(\theta_i - b_j))}{1 + \exp(a_j(\theta_i - b_j))} - \frac{1}{2}\right) + \varepsilon. \tag{8}$$

The resulting function in Equation (8) now *passes* through the point (b_j, ε), with new horizontal asymptotes $+0.5\delta + \varepsilon$ and $-0.5\delta + \varepsilon$.

In order to make the above function in Equation (8) bounded on [0, 1], values for δ and ε must be found so the function equals 0 when $\theta = 0$ and the function equals 1 when $\theta = 1$. Substituting accordingly yields the system of *two equations*:

$$\begin{cases} \delta\left(\dfrac{\exp(-a_j b_j)}{1 + \exp(-a_j b_j)} - \dfrac{1}{2}\right) + \varepsilon = 0 \\[2ex] \delta\left(\dfrac{\exp(a_j(1 - b_j))}{1 + \exp(a_j(1 - b_j))} - \dfrac{1}{2}\right) + \varepsilon = 1 \end{cases} \tag{9}$$

with *unknowns* δ and ε. Solving this system results in the following:

$$\begin{aligned} \delta &= \frac{(1 + \exp(a_j(1 - b_j)))(1 + \exp(a_j b_j))}{\exp(a_j) - 1}, \\[1ex] \varepsilon &= \frac{1}{2}\delta - \frac{(1 + \exp(a_j(1 - b_j)))}{\exp(a_j) - 1}. \end{aligned} \tag{10}$$

After substituting (10) in for δ and ε into the function in Equation (8), the function can be simplified as:

$$\frac{1 - \exp(a_j \times \theta_i)}{1 + \exp(a_j(\theta_i - b_j))} \times \frac{1 + \exp(a_j(1 - b_j))}{1 - \exp(a_j)}. \tag{11}$$

The resulting function (11) is a monotonically increasing link function with respect to θ_i with [0, 1] support. Finally, paralleling the 3PL IRTM in Equation (2), a pseudo-guessing parameter γ_j may be incorporated back into the function, yielding the desired three-parameter cognitive trait model (3P CTM):

$$\Pr(y_{ij} = 1 | \theta_i, \alpha_j, \beta_j, \gamma_j) = \gamma_j + (1 - \gamma_j) \times \frac{1 - \exp(\alpha_j \times \theta_i)}{1 + \exp(\alpha_j(\theta_i - \beta_j))} \times \frac{1 + \exp(\alpha_j(1 - \beta_j))}{1 - \exp(\alpha_j)}. \tag{12}$$

In the 3P CTM, I is the person index ($i = 1, \ldots, n$), j is the item index ($j = 1, \ldots, J$), α_j is a CTM slope or discrimination parameter (similar to the a parameter in the traditional model), β_j is a CTM location (or difficulty) parameter (similar to the b parameter in the traditional 2PL and 3PL IRTMs), and γ_j is a CTM pseudo-guessing parameter for item j.

Further parallels to IRTMs may also be derived. Setting $\gamma_j = 0$ yields a two-parameter cognitive trait model (2P CTM):

$$\Pr(y_{ij} = 1 | \theta_i, \alpha_j, \beta_j) = \frac{1 - \exp(\alpha_j \times \theta_i)}{1 + \exp(\alpha_j(\theta_i - \beta_j))} \times \frac{1 + \exp(\alpha_j(1 - \beta_j))}{1 - \exp(\alpha_j)}. \tag{13}$$

In addition, setting α_j to an arbitrary positive value, such as $\alpha_j = 10$, yields a one-parameter cognitive trait model (1P CTM):

$$\Pr(y_{ij} = 1|\theta_i, \beta_j) = \frac{1 - \exp(10 \times \theta_i)}{1 + \exp(10 \times (\theta_i - \beta_j))} \times \frac{1 + \exp(10 \times (1 - \beta_j))}{1 - \exp(10)}. \qquad (14)$$

For someone who wants the value fitted to *Beta*(2, 2) cdf instead of 10, use 9.3827 which is estimated with *fminsearch* in MATLAB (2020). All of the 1, 2, and 3P CTMs are not undefined when $\alpha_j = 0$. $\Pr(y_{ij} = 1|\theta_i, \alpha_j, \beta_j, \gamma_j)$ in 3P CTM Equation (12) converges to $\theta_i + \gamma_j$, and $\Pr(y_{ij} = 1|\theta_i, \alpha_j, \beta_j)$ in 2P CTM Equation (13) converges to θ_i as α_j approaches to 0, according to L'Hôpital's rule (e.g., Larson et al., 1999).

Figure 1 depicts a selection of 2P CTM Task Characteristic Curves (Item Characteristic Curves (ICCs) in the context of IRTMs) based on Equation (13). One of the well-known and attractive features of IRTMs is that the item location and person parameters can be presented on the same scale. For the proposed CTMs, the difficulty of an item also represents a location along the latent continuum, in this case from 0 to 1. As such, $\beta_j = 0$ represents the lowest location for item *j*, that is, the easiest item that everyone is expected to answer correctly, while $\beta_j = 1$ represents the highest location for item *j*, that is, the most difficult item, expected to be correctly answered only by a person at the mastery level.
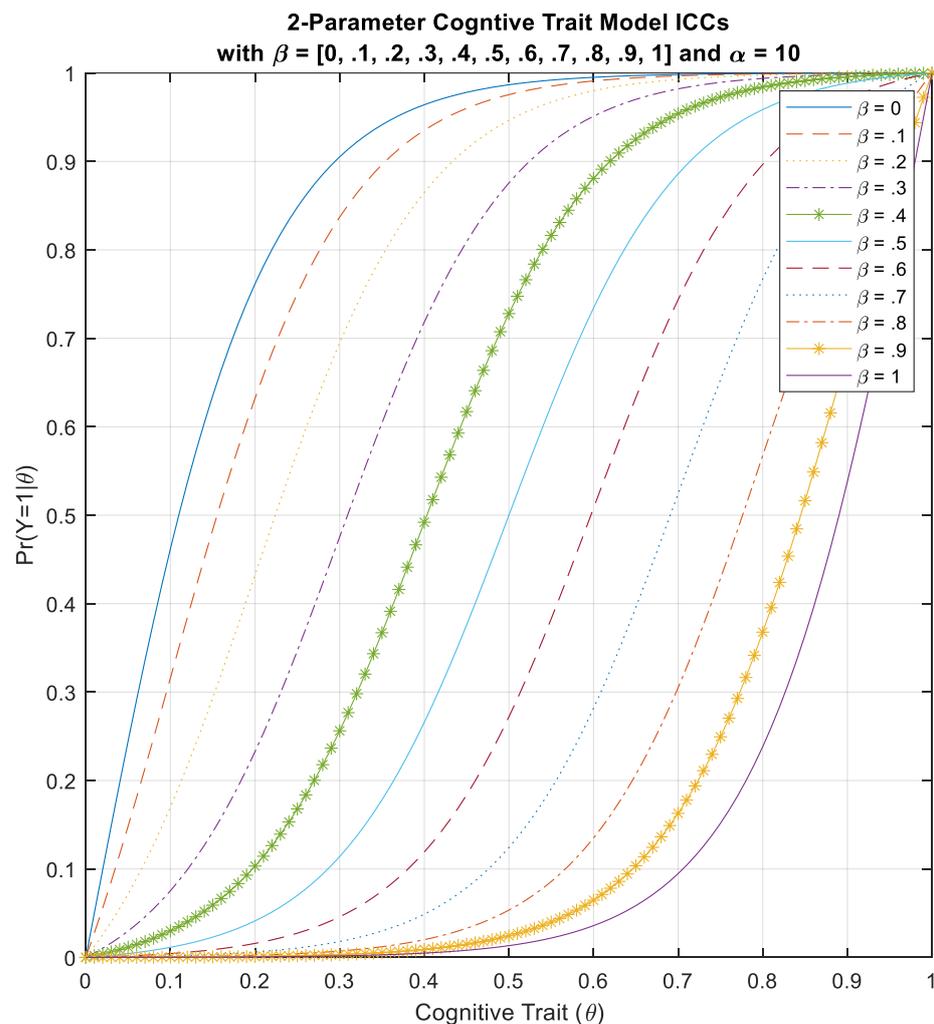
Figure 1. *Cont.*
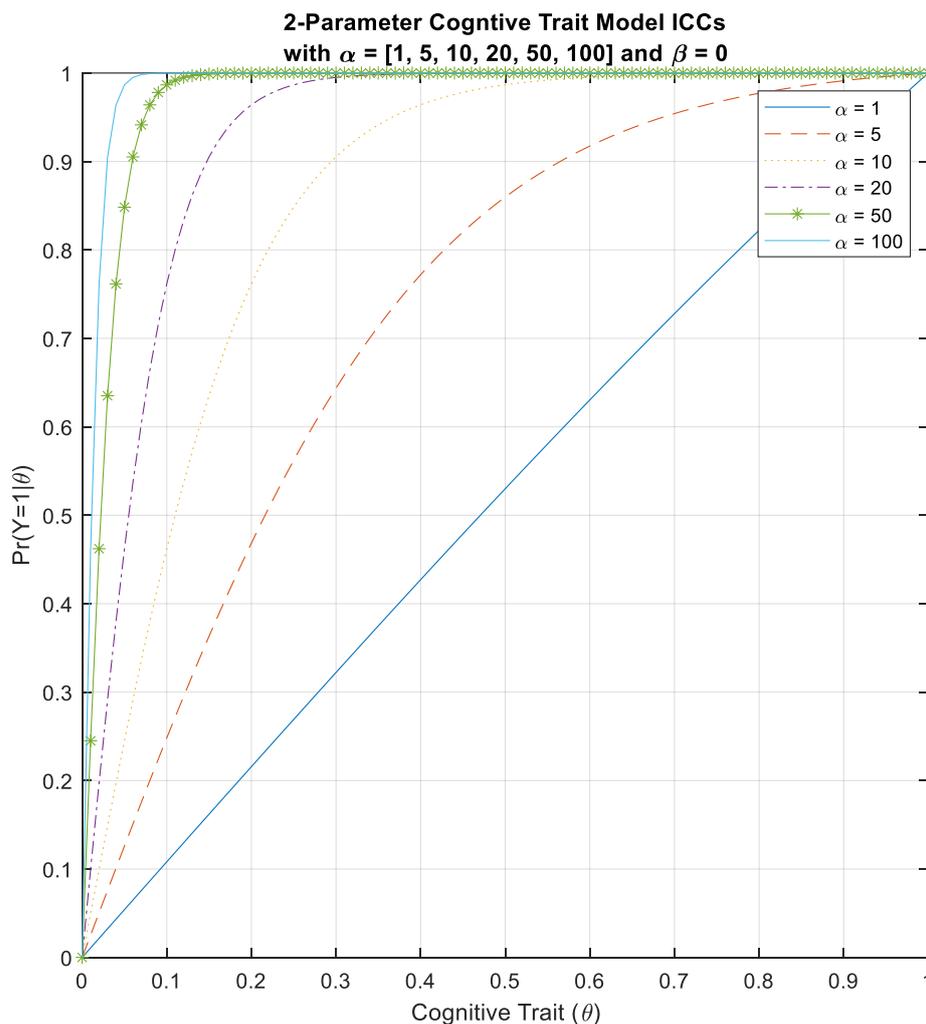
Figure 1. The top panel illustrates 2-Parameter Cognitive Trait Model Item Characteristic Curves (2P CTM ICCs) over different β values (0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1) when $\alpha = 10$. The bottom panel illustrates 2P CTM ICCs over different α values (1, 5, 10, 20, 50, 100) when $\beta = 0$.

The second technical facet is the CTM discrimination parameter α_j, which describes how well an item can differentiate between the people with traits below the item location and those with traits above the item location. Figure 2 depicts 2P CTM link function based on Equation (13), showing the different item discriminations. As with the traditional IRTMs, the discrimination parameter of the proposed model also reflects the steepness of the link function, such that the greater the value of α_j, the steeper the curve, and hence the better the item can discriminate at that location along the θ continuum. Conversely, the lower the value of α_j, the flatter the curve, and hence the less effectively the item can discriminate, because the probability of a correct response at low trait levels is similar to that at high trait levels.
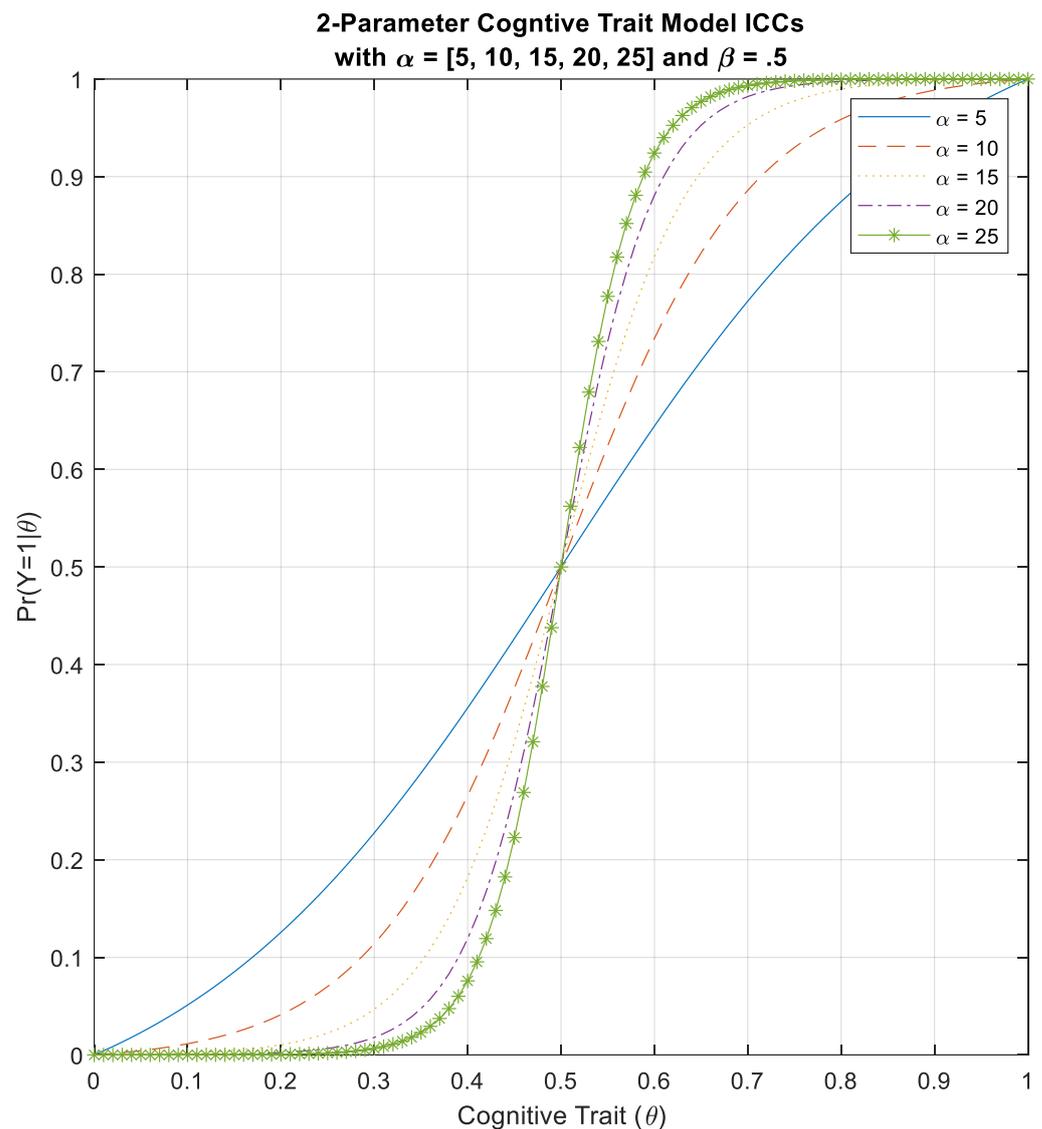
Figure 2. 2-Parameter Cognitive Trait Model Item Characteristic Curves over different α values (5, 10, 15, 20, 25) when β = 0.5.

The third component of the link function is the pseudo-guessing parameter γ_j, which conventionally characterizes the probability of the least capable person obtaining the correct answer due to guessing. Figure 3 depicts several 3P CTM link functions, based on Equation (12), constructed to show the different pseudo-guessing parameters. In the traditional 3PL models, this parameter represents the height of the lower asymptote of an ICC, and it shows the probability of $y_{ij} = 1$ for a person with a very low (lower asymptote) trait score. For the 3P CTM, however, γ_j represents the probability of getting a correct response when a person is *at* the lowest boundary, $\Pr(y_{ij} = 1 | \theta_i = 0)$, without introducing a low-left asymptote concept. Table 1 summarizes the types and interpretations of the scores over the different models.
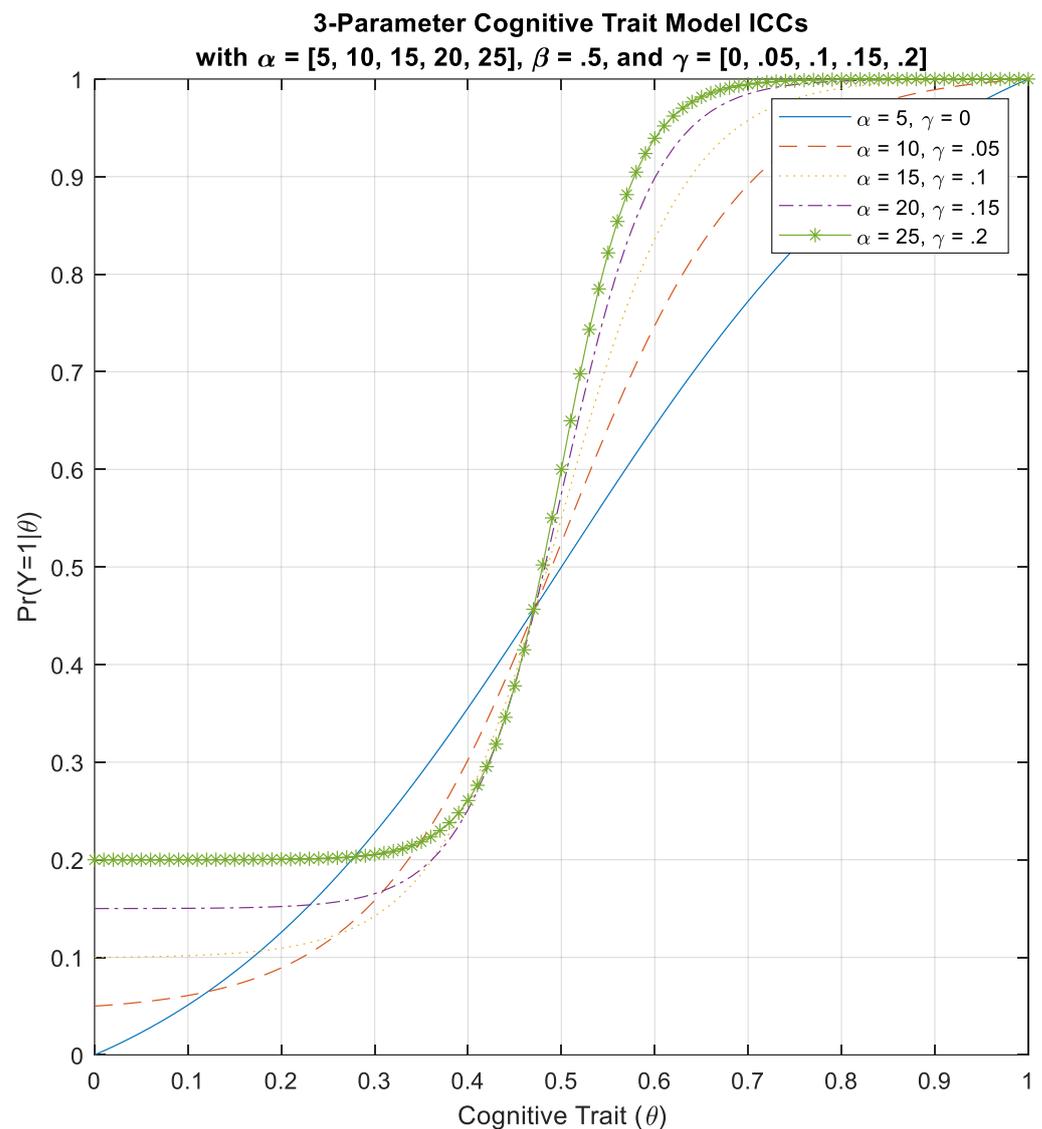
Figure 3. 3-Parameter Cognitive Trait Model Item Characteristic Curves over different *α* values (5, 10, 15, 20, 25) and *γ* values (0, 0.05, 0.1, 0.15, 0.2) when *β* = 0.5.

Table 1. Trait Level Interpretations over Different Measurement Models.

	Classical Test Theory Model	Diagnostic Classification Model	Logistic IRT Model and Factor Model	Cognitive Trait Model
Type	Continuous	Discrete	Continuous	Continuous
Range	0 to total score	0 or 1	$(-\infty, +\infty)$	[0, 1]
Interpretation of 0	All answers were wrong	No-mastery status	Mean of trait *	Ignorance level: The level of trait at which none of the tasks are expected to be successfully performed.
Interpretation of 0.5	Half of answers were correct	50% chance of mastery status **	Half *SD* above mean of trait *	Half of the mastery: The level of trait at which 50% of the tasks are expected to be successfully performed
Interpretation of 1	All answers were correct	Mastery status	One *SD* above mean of trait *	Mastery level; The level of trait at which all of the tasks are expected to be successfully performed

* 50% chance of answering correctly for the item difficulty value equal to the score 0 for 1, 2PL IRTMs; ** Probability of mastery status.

2.3. Likelihood Function and Markov Chain Monte Carlo Estimator

In the same way as those used in the traditional models, the likelihood function of the proposed models can be defined analytically without great difficulty. Furthermore, by virtue of previous work developing the estimator for the traditional model, it can be relatively straightforward to develop the Marginal Maximum Likelihood (MML; ref. [18]) estimator and other Bayesian estimates (expected a posteriori [EAP], maximum a posteriori [MAP]; refs. [17,19,20]), and MCMC for the proposed models.

During the last few decades, the MCMC estimation option has become an explosively popular technique for estimating a variety of statistical models, including IRTMs. Most importantly, this simulation-based estimation option enables one to simultaneously estimate both the items' and the subjects' parameters. Recently, in the context of IRTMs, this MCMC estimation has proved to be useful by providing great flexibility in dealing with a variety of modeling situations, including polytomous responses [21], nominal responses [22], missing data [21], hierarchical models [23], and so forth. For a more thorough treatments of the MCMC estimation in other psychometric models, see [24].

Beyond these modeling flexibilities, most importantly, an MCMC estimator can *simultaneously* estimate both the item and the person parameters, so that all of the parameter estimates appropriately reflect the estimation uncertainties of the other parameters. Furthermore, the MCMC estimation for the proposed models is as straightforward as the traditional model cases, as will be shown using the general purpose and freely available MCMC software, WinBUGS [25] to estimate both the person and the item parameters for the proposed models.

2.4. MCMC Specifications

The joint density specification for MCMC estimation of a 3P CTM can be expressed as:

$$\prod_{i=1}^{n}\prod_{j=1}^{J}\Pr(y_{ij}|\theta_i,\alpha_j,\beta_j,\gamma_j)\Pr(\theta_i|a_\theta^*,b_\theta^*)\Pr(\alpha_j|\mu_\alpha,\sigma_\alpha^2)\Pr(\beta_j|a_\beta^*,b_\beta^*)\Pr(\gamma_j|a_\gamma^*,b_\gamma^*), \tag{15}$$

where

$$y_{ij}|\theta_i,\alpha_j,\beta_j,\gamma_j \sim Bern[\Pr(y_{ij}=1|\theta_i,\alpha_j,\beta_j,\gamma_j)], \tag{16}$$

$$\Pr(y_{ij}=1|\theta_i,\alpha_j,\beta_j,\gamma_j) = \gamma_j + (1-\gamma_j) \times \frac{1-\exp(\alpha_j \times \theta_i)}{1+\exp(\alpha_j(\theta_i-\beta_j))} \times \frac{1+\exp(\alpha_j(1-\beta_j))}{1-\exp(\alpha_j)}, \tag{17}$$

$$\theta_i \sim Beta(a_\theta^*,b_\theta^*), \tag{18}$$

$$\alpha_j \sim N^+(\mu_\alpha,\sigma_\alpha^2) \text{ or } LN(\mu_a,\sigma_a^2), \tag{19}$$

$$\beta_j \sim Beta(a_\beta^*,b_\beta^*), \text{ and} \tag{20}$$

$$\gamma_j \sim Beta(a_\gamma^*,b_\gamma^*). \tag{21}$$

Several aspects of the above specification are worth addressing. First, in contrast with the traditional IRT model, the proposed models assume that the latent variable follows a Beta distribution, with two hyper-parameters a_θ^* and b_θ^*. These two parameters can be determined empirically or purposefully by a researcher. Note that, while the standard Beta distribution with a [0, 1] support interval can be easily generalized to an arbitrary interval using a linear transformation, the standard Beta distribution is adapted in this paper for modeling the person and the location parameter specifically with a [0, 1] interval. Note also that the traditional models (e.g., 2PL) would allow the discrimination parameter to be negative for fitting items with incorrectly scored. However, the CTM α parameter settings in Equation (21) only allow positive slopes. Unlike the traditional IRTMs, the CTMs only allow a positive slope, and as such cannot detect the aberrant items that are negatively related to the construct.

Many practical applications of the traditional models specify arbitrary values for the hyper-parameters to resolve the location and scale indeterminacies; the most popular choice is a normal mean = 0 and variance = 1 for the MML or MCMC estimators. Likewise, the CTMs also require one to specify values for a_θ^* and b_θ^*, such as $a_\theta^* = 2$ and $b_\theta^* = 2$. If these hyper-parameters are unknown, these can also be handled by hyper-prior distribution(s). For the discrimination parameters in (21), recognize that α_j is often used to specify a prior distribution that is restricted to the positive real line either with N^+ (a normal distribution truncated to the positive real line distribution) or with LN (a log-normal distribution). Again, the hyper-parameters μ_α and σ_α^2 are determined by the researcher, and the distributional specification of the discrimination parameter for the proposed models is the same as that of the traditional IRTMs. While the traditional models assume an unbounded continuum for the difficulty or location parameter, for the proposed model β_j, it is continuous but is bounded (from 0 to 1). Therefore, as seen in Equation (17), it is natural to adapt a [0, 1] continuous distribution, such as a standard Beta distribution, where a_β^* and b_β^* are the hyper-parameters that can be specified by the researcher. Therefore, similar to the traditional models, the difficulty and the person parameters are also located on the same [0, 1] continuum for the proposed models. Finally, in the traditional models, the pseudo-guessing parameters are the lower asymptote parameters and reflect the probability of a correct answer on the item when the proficiency is very low. For the proposed models, the pseudo-guessing parameters γ_j are also bounded by 0 and 1, and the Beta distribution can be the natural choice, as seen in Equation (21) with the two hyper-parameters, a_γ^* and a_γ^*. The next section of the paper illustrates the simultaneous estimation of the person and item parameters for the CTMs, using the MCMC estimator within the WinBUGS software [25].

3. Illustrations

3.1. Illustration with a Simulated Dataset

In this part of the illustration, I analyzed a simulated item responses dataset to evaluate the parameter replicability of the proposed models. In other words, I investigated how well the proposed model parameters are replicated, using a simulated dataset generated from the known item and person parameters. The details of these analysis steps are as follows.

First, generate 1000 person parameters from *Beta*(2, 2) distribution. Second, using the generated person parameters and the 20 items parameter values in Table 1, compute the probability of getting the right answer for all of the items using Equation (13). Third, generate the 20 item response data for each person as:

$$\begin{cases} y_{ij} = 1 & \text{if } \Pr(y_{ij} = 1|\theta_i, \alpha_j, \beta_j) > Uni(0,1) \\ y_{ij} = 0 & \text{if } \Pr(y_{ij} = 1|\theta_i, \alpha_j, \beta_j) \leq Uni(0,1) \end{cases}, \tag{22}$$

where *Uni* (0, 1) is a uniform distribution ranges from 0 to 1. Then, fit the 2P CTM in Equation (13) with the following prior distribution specifications (assuming prior independence among the parameters):

$$\theta_i \sim Beta(2,2), \tag{23}$$

$$\begin{aligned} \alpha_j &\sim LN(7.5, 0.1), \\ \beta_j &\sim Beta(2,2). \end{aligned} \tag{24}$$

Identical to the empirical data illustration case, I used the same *Beta*(2, 2) for the prior distributions of both the θ and β parameters, and these prior distributions for the latent variable and item parameters. A log-normal distribution $LN(7.5, 0.1)$ is also used for each α_j prior distribution (Appendix A gives the WinBUGS code for this analysis). The model was run with three chains with different starting values which are determined by the software. The first 2000 iterations were discarded as burn-in. Each chain was run for 3000 more iterations after burn-in, yielding 9000 iterations for use in summarizing the marginal distributions for all of the parameters. Convergence under the MCMC estimation has a different meaning than under the ML methods: a Markov chain is considered converged

when it becomes stationary, and this can be examined visually with trace plots and the diagnostic statistics developed by Brooks–Gelman–Rubin (BGR; ref. [26]).

In the outputs, no chain gets stuck in certain areas of the parameter space for all of the parameters, and all three chains of each parameter are mixed together. In other words, in this simulation data analysis, no trace plot of Markov chain was found unstationary after the burn-in period. Another way to visually inspect the convergence is the BGR diagnostic statistic. This ANOVA-type diagnostic compares within- and among-chain variance, and WinBUGS displays the BGR diagnostic statistic (red-line) when the "bgr diag" button is pressed in the "Sample monitor tool". The values around 1 indicate convergence, with 1.1 considered an acceptable limit [26,27]. The BGR diagnostic statistic also showed satisfactory convergence as the ratio (red) curves for all of the parameters approach 1. The marginal densities for the parameters are mostly unimodal and asymmetric, although some are asymmetric. Due to the space limitations, these trace plots are not provided in this paper. The results in their entirety can be requested from the author of this paper.

Table 2 gives the summary statistics of the marginal distributions for the item parameters and the first twenty person parameters. Table 2 also includes the true values for the item and person parameters. I examined the following summary statistics regarding the parameter replicability for the 2P CTM with the simulation data. First, I examined the correlation between the true parameter values and the point (i.e., the median of the marginal density) estimates. Second, for evaluating the bias of point estimates, I employed mean bias (*MB*) and mean relative bias (*MRB*) as:

$$MB = \sum_{i=1}^{NP} [\hat{\rho}_i - \rho_i]/NP, \tag{25}$$

$$MRB = \sum_{i=1}^{NP} [(\hat{\rho}_i - \rho_i)/\rho_i]/NP, \tag{26}$$

where ρ_i was the true value of ith parameter; $\hat{\rho}_i$ was the ith parameter point estimate given the estimation method converged; and *NP* was the number of parameters. Third, for evaluating the point estimates, I examined the root mean squared error (*RMSE*) with the following definitions:

$$RMSE = \left[\sum_{i=1}^{NP} (\hat{\rho}_i - \rho_i)^2/NP\right]^{1/2}. \tag{27}$$

Table 2. Simulation Data Marginal Density Summary Statistics.

Para.	True	Mean	Median	*SD*	Para.	True	Mean	Median	*SD*	Para.	True	Mean	Median	*SD*
α_1	5.000	5.166	5.083	0.600	β_1	0.100	0.107	0.107	0.060	θ_1	0.320	0.287	0.286	0.072
α_2	5.000	4.784	4.717	0.704	β_2	0.200	0.163	0.172	0.071	θ_2	0.281	0.270	0.268	0.069
α_3	5.000	4.316	4.289	0.753	β_3	0.300	0.211	0.227	0.079	θ_3	0.133	0.157	0.152	0.060
α_4	5.000	4.105	4.137	0.837	β_4	0.400	0.302	0.322	0.077	θ_4	0.814	0.729	0.732	0.067
α_5	5.000	3.674	3.873	1.221	β_5	0.500	0.544	0.533	0.083	θ_5	0.261	0.233	0.230	0.070
α_6	5.000	3.389	3.438	0.927	β_6	0.600	0.667	0.636	0.097	θ_6	0.429	0.506	0.508	0.075
α_7	5.000	3.465	3.376	0.541	β_7	0.700	0.856	0.857	0.081	θ_7	0.109	0.123	0.118	0.056
α_8	5.000	4.690	4.662	0.759	β_8	0.800	0.796	0.781	0.073	θ_8	0.439	0.411	0.409	0.075
α_9	5.000	4.736	4.690	0.373	β_9	0.900	0.944	0.953	0.042	θ_9	0.566	0.662	0.663	0.071
α_{10}	5.000	5.480	5.413	0.463	β_{10}	1.000	0.937	0.944	0.044	θ_{10}	0.419	0.321	0.319	0.073
α_{11}	10.000	9.890	9.733	1.060	β_{11}	0.100	0.063	0.062	0.037	θ_{11}	0.344	0.352	0.349	0.076
α_{12}	10.000	11.740	11.680	1.458	β_{12}	0.200	0.187	0.190	0.025	θ_{12}	0.641	0.755	0.758	0.068
α_{13}	10.000	10.120	10.090	0.968	β_{13}	0.300	0.304	0.305	0.016	θ_{13}	0.764	0.761	0.765	0.064
α_{14}	10.000	10.790	10.760	0.962	β_{14}	0.400	0.391	0.391	0.013	θ_{14}	0.863	0.862	0.867	0.055
α_{15}	10.000	9.271	9.262	0.826	β_{15}	0.500	0.480	0.480	0.013	θ_{15}	0.935	0.893	0.899	0.051
α_{16}	10.000	10.940	10.910	0.969	β_{16}	0.600	0.600	0.600	0.012	θ_{16}	0.748	0.762	0.764	0.065
α_{17}	10.000	9.936	9.911	1.017	β_{17}	0.700	0.721	0.720	0.019	θ_{17}	0.637	0.615	0.616	0.075
α_{18}	10.000	9.509	9.487	1.147	β_{18}	0.800	0.831	0.825	0.034	θ_{18}	0.492	0.542	0.543	0.076
α_{19}	10.000	10.600	10.500	1.352	β_{19}	0.900	0.900	0.895	0.040	θ_{19}	0.303	0.316	0.314	0.073
α_{20}	10.000	12.460	12.250	1.358	β_{20}	1.000	0.957	0.961	0.029	θ_{20}	0.642	0.570	0.571	0.076

As seen in Table 3, the *MB* values are low (i.e., less than 0.1) for all three parameters. The *MB* values for the β and θ parameters are especially low (i.e., less than 0.01). All three of the parameters' *MRB* values are also low (i.e., less than 0.05; that is, the bias size relative to a parameter value is less than 5%). The β and θ parameters' *RMSE* values are also low (i.e., less than 0.05). The *RMSE* values (i.e., 0.962; more than 0.05) of the α parameter shows the parameter estimation is relatively unstable than the β and θ parameters. It is recommended that readers evaluate the proposed model by noting the instability of the α parameter estimation. Lastly, the correlations between the true parameter values and the estimates are high (i.e., greater than 0.947) for all three parameters. In Figure 4, the scatter plots of true person trait values versus the 2P CTM estimates are shown, that also represent a linear relationship between the true values and the estimates.

Table 3. Statistics for the Parameter Replicability Analysis.

Parameter	NP	MB	MRB	RMSE	Pearson's r
α	20	-0.087	-0.040	0.962	0.968
β	20	-0.002	-0.032	0.050	0.987
θ	1000	-0.004	0.021	0.071	0.947
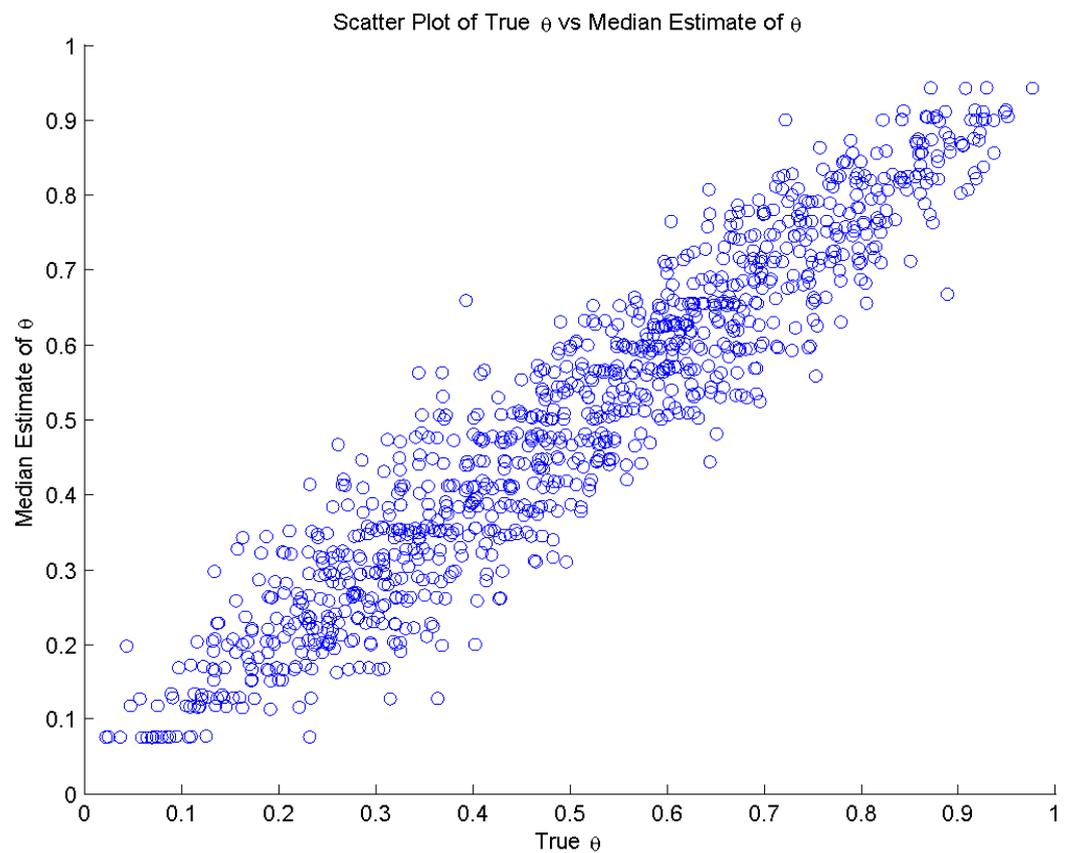
Figure 4. Scatter Plot of True and 2P CTM Median Estimate of θ. This figure illustrates the relationship between true and 2P CTM median estimates from the simulation dataset ($n = 1000$).

Table 4 also provides several interesting aspects of the person parameter estimates among the different models. First, θ_{True} is highly correlated with all of the other three estimates with similar degrees (i.e., around 0.94 coefficients). Second, the Spearman's rho coefficient between $\hat{\theta}_{2PL}$ and $\hat{\theta}_{2PTL}$ is perfect as 1. In other words, the rank order between the two model estimates is exactly the same in the simulated data case (especially for these 20 items). Third, $\hat{\theta}_{2PTL}$ is highly correlated (i.e., >0.99) with $\hat{\theta}_{CP}$, but the correlation coefficients

are less than those between $\hat{\theta}_{2PL}$ and $\hat{\theta}_{2PTL}$. That means, although $\hat{\theta}_{CP}$ and $\hat{\theta}_{2PL}$ are both a score from 0 to 1 and highly correlated, $\hat{\theta}_{2PL}$ and $\hat{\theta}_{2PTL}$ are more statistically similar.

Table 4. Correlation Table of the Simulation Data Person Parameter Estimates.

		θ_{True}	$\hat{\theta}_{CP}$	$\hat{\theta}_{2PL}$	$\hat{\theta}_{2PTL}$
θ_{True}	Pearson's *r*	1.000	0.944 **	0.944 **	0.947 **
	Spearman's *rho*	1.000	0.945 **	0.947 **	0.947 **
$\hat{\theta}_{CP}$	Pearson's *r*	0.944 **	1.000	0.993 **	0.995 **
	Spearman's *rho*	0.945 **	1.000	0.996 **	0.995 **
$\hat{\theta}_{2PL}$	Pearson's *r*	0.944 **	0.993 **	1.000	0.997 **
	Spearman's *rho*	0.947 **	0.996 **	1.000	1.000 **
$\hat{\theta}_{2PTL}$	Pearson's *r*	0.947 **	0.995 **	0.997 **	1.000
	Spearman's *rho*	0.947 **	0.995 **	1.000 **	1.000

** Correlation is significant at the 0.01 level (2-tailed); θ_{True} is the true trait given 2P CTM is correct model; $\hat{\theta}_{CP}$ is the correct answer ratio (i.e., the number of correctly answered items/the total number of items) estimate; $\hat{\theta}_{2PL}$ is 2PL estimate; $\hat{\theta}_{2PTL}$ is 2P CTM estimate.

The interpretation of the high correlations between the parameters of the two model frameworks is as follows. CTM's link function was analytically derived by transforming IRTM's link function, the logistic function. Therefore, it is quite natural that the parameters of the two models have similar properties and are highly correlated. However, because the individual parameter in CTM has a certain value in the 0 to 1 bounded continuum, the 'conceptualization' and 'interpretation' of the parameter are different. Thus, the usage would be different for the different purposes of the assessments, such as summative vs. formative assessments. These different usages will be described in more detail in the discussion section.

Note that both $\hat{\theta}_{2PL}$ and $\hat{\theta}_{2PTL}$ are based on LVMs, and $\hat{\theta}_{CP}$ is an average of the observed variables. For example, in the simulation dataset, the 809th and 231st subjects' $\hat{\theta}_{CP}$ are both 0.8 (80% correct). However, the $\hat{\theta}_{2PTL}$ are different as 0.8064 (interpreted as 80.640% of the mastery level, or the trait level at which 80.64% of the tasks are expected to correctly be answered) for the 809th subject with a response vector {1, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0} and 0.7741 (interpreted as 77.41% of the mastery level, or the trait level at which 77.41% of the tasks are expected to correctly answered) for the 231st subject with a response vector {1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0}. This means that, although the averages of the observed variables ($\hat{\theta}_{CP}$) are the same, the LVM-based estimates ($\hat{\theta}_{2PTL}$) can be different, because $\hat{\theta}_{2PTL}$ is taking into account the different characteristics (i.e., difficulties and/or discriminations) of the items (see Table 2 for the population item difficulty parameter values). For the 211st subject with a response vector {1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0}, $\hat{\theta}_{2PL} = -1.284$ and $\hat{\theta}_{2PTL} = 0.2037$. The $\hat{\theta}_{2PL}$ value of -1.284 provides a relative standing information as "1.284 *SD* below than average". However, the $\hat{\theta}_{2PTL}$ value of 0.2037 can be interpreted as "20.37% of the mastery level, or the trait level at which 20.37% of the tasks are expected to be correctly answered". In sum, this simulated dataset analysis demonstrates that it is possible not only to satisfactorily estimate all of the item and person parameters, but also to successfully replicate the proposed 2P CTM parameters.

3.2. Illustration with an Empirical Dataset

To illustrate the proposed models, I analyzed the item responses data originally provided by [28]. These were responses from 1000 examinees to five dichotomously scored items on the Section 6 of the Law School Admissions Test (LSAT). There are two reasons

for choosing this example as follows. First of all, this example has been used in a wide variety of LVMs, especially IRTM studies. Second, all of the items were in the multiple choice format with five response options. Therefore, the 3P CTM in Equation (12) can be demonstrated with this dataset.

For the MCMC estimation, the model was fitted with the following prior distribution specifications (assuming prior independence among the parameters):

$$\theta_i \sim Beta(2,2), \tag{28}$$

$$\alpha_j \sim LN(5,10), \tag{29}$$

$$\beta_j \sim Beta(2,2), \text{ and} \tag{30}$$

$$\gamma_j \sim Beta(7,25). \tag{31}$$

The same $Beta(2,2)$ is used for both the θ and β parameters' prior distributions, to reflect a moderately strong and symmetric belief on the parameters around 0.5. Obviously, other a priori knowledge and empirical evidences of the θ and β distributions can be incorporated by choosing different values of the shape parameters. $Beta(2,2)$ is symmetric with 0.5 mean, variance as $(\alpha^* \times \beta^*)/\left\{(\alpha^* + \beta^*)^2 \times (\alpha^* + \beta^* + 1)\right\} = 4/80 = 0.05$, and it is flatter than the normal distributions (i.e., less informative or relatively diffuse than the normal prior). Given the assumption that the chance of having a correct response increases monotonically with latent trait, using a log-normal normal distribution, $\alpha_j \sim LN(5,10)$, which is also relatively diffuse, ensures that each $\alpha_j > 0$.

Because the assessment is a five-response option multiple choice format, as a moderately informative prior out of several distributions, $Beta(7,25)$ has been adapted as the prior distribution of γ. $Beta(7,25)$ mode is $(\alpha^* - 1)/(\alpha^* + \beta^* - 2) = 6/30 = 0.2$, mean is $\alpha^*/(\alpha^* + \beta^*) = 7/32 = 0.21875$, and the variance is $(\alpha^* \times \beta^*)/\left\{(\alpha^* + \beta^*)^2 \times (\alpha^* + \beta^* + 1)\right\} = 175/33792 \approx 0.00517$. Note that $Beta(7,25)$ is a Beta distribution for reflecting the five choices item pseudo-guessing parameter (i.e., 1/Number of Choices = 1/5 = 0.2). Other Beta distributions could be adapted to model the multiple choice items, e.g., Beta distributions with mode as $(\alpha^* - 1)/(\alpha^* + \beta^* - 2) = 1/$Number of Choices. Note that the sensitivity analysis of the prior selection on the estimates would be an important issue in MCMC. In general, the prior sensitivity is a function of the data size, i.e., a prior selection will be more influential to the estimates as the data size gets smaller.

The WinBUGS code for this analysis is given in Appendix A. The model was run with five chains. The first 1000 iterations were discarded as burn-in, and each chain was run for 1000 more iterations after burn-in, yielding 5000 iterations for use in summarizing the marginal distributions for all of the parameters. The trace plot (i.e., chain history) is a plot of the MCMC iteration number against the value of the draw of the parameter at each iteration. In the chain history, no chain gets stuck in certain areas of the parameter space, and all five chains of each parameter are mixed together. In this dataset result, the ratio (red) curves for all of the parameters approach to 1. In the MCMC outputs, the marginal densities are mostly unimodal and asymmetric, with some of them are asymmetric. Due to space limitations, the chain history and BGR diagnostic plots are not provided in this manuscript. The results in their entirety can be requested from the author of this paper. Accordingly, Table 5 contains the summary statistics of the marginal distributions for the item parameters and the first five-person parameters. The overall analysis of the results not only shows that the 3P CTM (all of the item and person parameters) can be estimated satisfactorily with the LSAT data, but also shows that one can straightforwardly analyze the proposed models via the MCMC estimation option with conventional MCMC software, such as WinBUGS (Appendix A gives the code for this analysis).

Table 5. LSAT Data Marginal Density Summary Statistics.

	3PL IRTM					3P CTM				
Parameter	Mean	Median	SD	2.5% Perc.	97.5% Perc.	Parameter	Mean	Median	SD	2.5% Perc.	97.5% Perc.
A_1	1.08	1.03	0.31	0.65	1.83	α_1	8.86	8.60	8.60	1.70	6.52
a_2	0.96	0.88	0.42	0.47	2.04	α_2	1.25	0.95	0.95	1.17	0.02
a_3	1.41	1.26	0.65	0.62	3.26	α_3	2.61	2.18	2.18	2.44	0.06
a_4	0.84	0.79	0.29	0.46	1.54	α_4	1.98	1.96	1.96	1.42	0.03
a_5	0.85	0.83	0.22	0.49	1.36	α_5	5.21	5.19	5.19	1.03	3.33
b_1	−2.50	−2.45	0.52	−3.64	−1.59	β_1	0.08	0.08	0.08	0.04	0.01
b_2	−0.58	−0.61	0.43	−1.38	0.30	β_2	0.43	0.42	0.42	0.21	0.08
b_3	0.34	0.33	0.24	−0.11	0.84	β_3	0.67	0.70	0.70	0.18	0.21
b_4	−1.10	−1.09	0.42	−1.94	−0.27	β_4	0.35	0.32	0.32	0.20	0.05
b_5	−2.13	−2.06	0.51	−3.29	−1.27	β_5	0.13	0.11	0.11	0.08	0.02
c_1	0.26	0.25	0.10	0.10	0.47	γ_1	0.33	0.33	0.33	0.10	0.16
c_2	0.26	0.25	0.10	0.09	0.49	γ_2	0.38	0.39	0.39	0.05	0.26
c_3	0.23	0.22	0.08	0.09	0.37	γ_3	0.20	0.19	0.19	0.07	0.09
c_4	0.25	0.25	0.09	0.09	0.46	γ_4	0.45	0.47	0.47	0.07	0.29
c_5	0.27	0.26	0.10	0.10	0.48	γ_5	0.44	0.44	0.44	0.10	0.26
θ_1	−1.75	−1.75	0.75	−3.26	−0.26	θ_1	0.14	0.12	0.12	0.09	0.02
θ_2	−1.75	−1.76	0.75	−3.22	−0.26	θ_2	0.14	0.13	0.13	0.08	0.02
θ_3	−1.74	−1.72	0.74	−3.21	−0.30	θ_3	0.14	0.13	0.13	0.08	0.02
θ_4	−1.36	−1.35	0.77	−2.89	0.12	θ_4	0.19	0.17	0.17	0.11	0.03
θ_5	−1.36	−1.35	0.77	−2.91	0.12	θ_5	0.19	0.17	0.17	0.11	0.03

For the guessing parameters in Table 5, the CTM guessing parameters tend to be bigger than those from the PL model. This result also clearly shows the differences between the corresponding parameters of the two models. In the case of the PL model, the guessing parameter value (c_1 = 0.26) can be interpreted as the probability of getting the right answer for the first item is 0.26 for someone whose latent trait level is negatively very low (theoretically, it requires the low-left asymptote concept). In the case of the CTM, the CTM guessing parameter estimate (γ_1 = 0.33) can be straightforwardly interpreted as the probability of getting the right answer for the first item is 0.33 for a person at the lowest (a trait level for 0% chance of successfully completing any task) level.

Table 6 also provides the association measures of the person parameter estimates among the different models. Note that all of Spearman's rho coefficients between the IRTM and CTM estimates are high (i.e., >0.85), but those are not perfect 1 as in Table 4. In other words, the ranking information between the two models (i.e., 2PL IRTM and 2P CTM; 3PL IRTM and 3P CTM) are not exactly same for this empirical data simulation dataset. This difference would be due to the difference in the number of items (i.e., 20 items vs. 5 items). In sum, this simulated dataset analysis demonstrates that it is possible to satisfactorily estimate all of the 3P CTM parameters.

Table 6. Correlation Table of the LSAT Data Person Parameter Estimates.

		$\hat{\theta}_{2PL}$	$\hat{\theta}_{2PTL}$	$\hat{\theta}_{3PL}$	$\hat{\theta}_{3PTL}$
$\hat{\theta}_{2PL}$	Pearson's r	1.000	0.953 **	0.985 **	0.940 **
	Spearman's rho	1.000	0.869 **	0.948 **	0.922 **
$\hat{\theta}_{2PTL}$	Pearson's r	0.953 **	1.000	0.987 **	0.986 **
	Spearman's rho	0.869 **	1.000	0.896 **	0.918 **
$\hat{\theta}_{3PL}$	Pearson's r	0.985 **	0.987 **	1.000	0.980 **
	Spearman's rho	0.948 **	0.896 **	1.000	0.957 **
$\hat{\theta}_{3PTL}$	Pearson's r	0.940 **	0.986 **	0.980 **	1.000
	Spearman's rho	0.922 **	0.918 **	0.957 **	1.000

** Correlation is significant at the 0.01 level (2-tailed).

4. Discussion

This paper makes several contributions to psychometrics. First, this paper proposes a perspective of operationalizing the quantity of a cognitive domain trait (such as knowledge, skill, ability, expertise, intelligence, achievement, competence, and proficiency) as a level between two boundaries. (i.e., the lowest ignorance level and the highest mastery level). This perspective enables us to conceptualize and interpret the trait levels as a relative magnitude to the mastery level, instead of relative standings on an unbounded continuum. The perspective makes us rethink the question of "relative standings or magnitudes?" on cognitive trait levels and provides an option to cope with it. The placement of a person parameter's boundaries as 0 and 1 allows us to conceptualize a person's ability level, based on the construction's 0 (ability level where no tasks can be performed correctly) and 1 (ability level where all tasks are performed correctly). Thus, such conceptualization of the person parameter value facilitates the usages and interpretations in domain mastery and/or learning growth over time/stimulus of a person. This new framework will be able to better respond to the demands of digital assessments, which require th e interpretation and analysis of individual growth and change beyond inter-individual comparisons, such as formative assessment, learning progression/growth, or adaptive learning systems.

Second, for the above point of view, this paper presents an analytical framework for cognitive domain traits: cognitive trait models (CTM) which have supports on the [0, 1] latent continuum using truncated logistic functions. After illustrating how the new model parameters can be estimated and replicated via the MCMC estimation method within a Bayesian framework, the second primary contribution becomes possible. Specifically, it is possible to simultaneously estimate both the item and the person parameters by making minor modifications of the MCMC specifications of the traditional model (two illustrations were exemplified, with program syntaxes provided). The implementation illustrations provide modeling accessibility to applied researchers who are interested in adopting the proposed models.

Third, due to the popularity of the logistic function family, the above contributions of this paper have great potential to be extended beyond psychometrics. Since initially introduced by [29], the logistic function and variants of it have proved useful for modeling a wide variety of phenomena in diverse research domains, such as population ecology [30], technological innovation infusion [31], biological and genetics [32], artificial neural networks [33,34], and nonlinear growth processes [35,36]. Due to such a wide range of applications of the logistic function family, the proposed framework with [0, 1] support may have great potential to provide theoretically and/or practically sound analytical option in diverse fields.

Note that, as articulated earlier, the link function of CTM is analytically derated from the IRTMs' logistic function. Therefore, the mathematical/analytical properties of the two model frameworks are similar. However, because the person parameters of CTM are bound to 0 and 1, the interpretations and usages are different from those of the IRTMs. The CTM is useful for criterion/domain-referenced and/or within-person comparisons, such as formative assessment, learning analytics, and learning progression model applications. In addition, the IRTMs, as is already well known, are useful for the use and interpretation of norm-referenced or summative assessment/testing, with the primary purpose of operationalizing the differences between the subjects. It is important to understand the difference between the two modeling frameworks, and accordingly to use them appropriately for the purposes of making relevant interpretations.

It should be noted that the proposed framework requires the clarity and stability of the mastery level, which is a cognitive level where all of the tasks are expected to be successfully performed. Therefore, the meanings and interpretations of the CTM model parameters are subject to the degree of identification and specifications of the domain of the tasks (as the traditional framework does). Note that the interpretations and inferences of the CTMs also require the assumption of "random samples from the domain of tasks", as the other traditional frameworks do. In the proposed framework, because the trait quantity

is a number from the 0 to 1 continuum, the uncertainties of the domain of tasks (especially in the top range) would impact the whole metric of a trait. Analytical investigations, adjustments, and treatments for such issues (for example, more difficult or easy tasks are added to the domain; equating or vertical scaling) remain as future research topics.

Although the applicability of the proposed models is imminent for the traits in the cognitive domain, it is not imminent for the traits in the affective domain (e.g., beliefs or attitudes; ref. [2]). In the affective domain, a negative estimate from the traditional models could be more useful to conceptualize or interpret. An attempt to integrate the proposed framework into the affective domain induces bounded affective continuums, such as from "completely not believe" to "completely believe", or from "completely disagree" to "completely agree". Such applications would require a substantial amount of theoretic considerations for each affective trait, and remains as future work for the experts.

In the last few decades, IRTMs [6] based on the logistic function with unbounded support have been popular measurement models in the educational and psychological domain. Although this study focuses on introducing a new conceptualization of the trait level itself, the proposed models can be considered as a type of IRTM, so called Truncated Logistic IRTMs (i.e., 1, 2, 3PTL IRTMs). From this point of view, the straightforward connection between the person and item difficulty parameters (i.e., when $\theta = \beta$, the probability of getting an item right is always 0.5) is lost in CTMs. In other words, the probability of getting an item right when $\theta = \beta$ depends on α in the CTMs. Considering the CTMs as a type of IRTM would lead to a need for IRTM-specific features and analyses, such as test characteristic curves (TCCs), item information functions (IIFs), test information functions (TIFs), standard error of measurement (SEM), or model-parameter invariance investigation, etc. [37,38]. In particular, the TCC and TIF for the CTMs will be useful improvements, because they allow us to investigate the relationship between observations and trait levels at the level of the task domain, not at the item level. Furthermore, a didactic presentation with a walkthrough of the modification on the original and the transformed sigmoid with other dataset examples would be beneficial to someone who would like to understand more details of the analytic characteristics of the CTM. Such extensions and articulations are left to future research.

Beyond the conceptualization and interpretability of the model parameters, other modeling aspects (e.g., the robustness of the parameter invariance, model fits, sample or item size requirements) should be further evaluated for someone who is considering the new models.

It would be interesting to investigate how the CTM trait levels are numerically different from the DCM or the Beta-Binomial model [9] estimates. Note that other traditional LVMs (such as IRTMs, DCMs, or FMs) and CTMs are not within a nested structure. Therefore, the comparisons of these aspects would require an intense series of investigations (e.g., Monte Carlo simulation studies varying several conditions: ground truth models for generating data, sample sizes, and/or number of tasks, etc.) which also remain subject to future work. Lastly, the traditional psychometric modeling options have been vigorously extended, including multi-latent traits, polytomous observables, longitudinal, and multi-level/nested structure, etc. [39]. As this paper has sought to demonstrate, one may implement these extensions for the proposed analytical framework by virtue of the Bayesian framework, which is flexible in dealing with a variety of modeling situations. Such developments are eagerly awaited.

Appendix A. WinBUGS Codes for 1P, 2P, and 3P Cognitive Trait Models

```
# 3P CTM
model
    {
        for (i in 1: n) {
            for (j in 1: J) {
                p[I, j] <- gamma[j] + (1-gamma[j])*(1-exp(alpha[j]*theta[i]))/(1+exp
(alpha[j]*(theta[i]-beta[j])))*(1+exp(alpha[j]*(1-beta[j])))/(1-exp(alpha[j]))
                Y[i, j] ~ dbern(p[I, j])
            }
            theta[i] ~ dbeta(2, 2)
        }
    # Priors
        for (j in 1:J) {
            alpha[j] ~ dlnorm(5, 0.1)
            beta[j] ~ dbeta(2, 2)
            gamma[j] ~dbeta(7,25)
        }
    }
# 2P CTM
model
    {
        for (i in 1: n) {
        for (j in 1: J) {
                p[i, j] <- (1-exp(alpha[j]*theta[i]))/(1+exp(alpha[j]*(theta[i]-beta[j])))
*(1+exp(alpha[j]*(1-beta[j])))/(1-exp(alpha[j]))
                Y[I, j] ~ dbern(p[i, j])
            }
            theta[i] ~ dbeta(2, 2)
        }
        for (j in 1:J) {
            alpha[j] ~ dlnorm(5, 0.1) # dlnorm(7.5, 1) for the Simulation Data
            beta[j] ~ dbeta(2, 2)
        }
    }
# 1P CTM
model
    {
        for (i in 1: n) {
        for (j in 1: J) {
                p[i, j] <- (1-exp(alpha[j]*theta[i]))/(1+exp(alpha[j]*(theta[i]-beta[j])))
*(1+exp(alpha[j]*(1-beta[j])))/(1-exp(alpha[j]))
                Y[I, j] ~ dbern(p[i, j])
            }
            theta[i] ~ dbeta(2, 2)
        }
        for (j in 1:J) {
            alpha[j] <- 10
            beta[j] ~ dbeta(2, 2)
        }
    }
```

References

1. Messick, S. Validity of test interpretation and use. *ETS Res. Rep. Ser.* **1990**, 1487–1495. [CrossRef]
2. Bloom, B.S.; Engelhart, M.D.; Furst, E.J.; Hill, W.H.; Krathwohl, D.R. (Eds.) *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*; David McKay Co. Inc.: New York, NY, USA, 1956.

3. American Educational Research Association; American Psychological Association; National Council on Measurement in Education; Joint Committee on Standards for Educational; Psychological Testing (US). *Standards for Educational and Psychological Testing*; American Educational Research Association: Washington, DC, USA, 2014.

4. Baker, F.B.; Kim, S.H. *Item Response Theory Parameter Estimation Techniques*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2004.

5. National Governors Association Center for Best Practices; Council of Chief State School Officers. Common Core State Standards for Mathematics: Kindergarten. 2010. Available online: http://www.corestandards.org/Math/Content/K (accessed on 1 January 2020).

6. Lord, F.M. *Applications of Item Response Theory to Practical Testing Problems*; Lawrence Erlbaum: Hillsdale, NJ, USA, 1980. [CrossRef]

7. Bock, R.D.; Thissen, D.; Zimowski, M.F. IRT estimation of domain scores. *J. Educ. Meas.* **1997**, *34*, 197–211. [CrossRef]

8. Wilson, M.; Draney, K. Standard Mapping: A technique for setting standards and maintaining them over time. Paper in an invited symposium: "Models and analyses for combining and calibrating items of different types over time". In Proceedings of the International Conference on Measurement and Multivariate Analysis, Banff, AB, Canada, 11–14 May 2000.

9. Wilcox, R.R. A Review of the Beta-Binomial Model and Its Extensions. *J. Educ. Stat.* **1981**, *6*, 3–32. [CrossRef]

10. Lee, W.-C.; Hanson, B.A.; Brennan, R.L. Estimating Consistency and Accuracy Indices for Multiple Classifications. *Appl. Psychol. Meas.* **2002**, *26*, 412–432. [CrossRef]

11. von Davier, M.; Lee, Y.-S. (Eds.) *Handbook of Diagnostic Classification Models*; Springer: New York, NY, USA, 2019.

12. Almond, R.; Dibello, L.V.; Moulder, B.; Zapata-Rivera, D. Modeling Diagnostic Assessments with Bayesian Networks. *J. Educ. Meas.* **2007**, *44*, 341–359. [CrossRef]

13. Gilks, W.R.; Richardson, S.; Spiegelhalter, D.J. *Introducing Markov Chain Monte Carlo*; Gilks, W.R., Richardson, S., Spiegelhalter, D.J., Eds.; Markov chain Monte Carlo in practice; Chapman and Hall: London, UK, 1996; pp. 1–19.

14. Birnbaum, A. *Estimation of an Ability*; Lord, F.M., Novick, M.R., Eds.; Statistical theories of mental test scores; Addison-Wesley: Reading, MA, USA, 1968; pp. 423–479.

15. Lord., F.M.; Novick, M.R. *Statistical Theories of Mental Test Scores*; Addison-Wesley: Reading, MA, USA, 1968.

16. Rasch, G. *Studies in Mathematical Psychology: I. Probabilistic Models for Some Intelligence and Attainment Tests*; Nielsen & Lydiche: Copenhagen, Denmark, 1960.

17. Bock, R.D.; Mislevy, R.J. Adaptive EAP Estimation of Ability in a Microcomputer Environment. *Appl. Psychol. Meas.* **1982**, *6*, 431–444. [CrossRef]

18. Bock, R.D.; Aitkin, M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* **1981**, *46*, 443–459. [CrossRef]

19. Choi, J.; Kim, S.; Chen, J.; Dannels, S. A Comparison of Maximum Likelihood and Bayesian Estimation for Polychoric Correlation Using Monte Carlo Simulation. *J. Educ. Behav. Stat.* **2011**, *36*, 523–549. [CrossRef]

20. Sorenson, H.W. *Parameter Estimation: Principles and Problems*; Marcel Dekker: New York, NY, USA, 1980.

21. Patz, R.; Junker, B.W. Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *J. Educ. Behav. Stat.* **1999**, *24*, 342–366. [CrossRef]

22. Wollack, J.A.; Bolt, D.M.; Cohen, A.S.; Lee, Y.S. Recovery of item parameters in the nominal response model: A comparison of Marginal Maximum Likelihood estimation and Markov Chain Monte Carle estimation. *Appl. Psychol. Meas.* **2002**, *26*, 339–352. [CrossRef]

23. Fox, J.-P.; Glas, C.A.W. Bayesian Estimation of a Multilevel IRT Model using Gibbs Sampling. *Psychometrika* **2001**, *66*, 269–286. [CrossRef]

24. Levy, R.; Mislevy, R.J. *Bayesian Psychometric Modeling*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2017.

25. Spiegelhalter, D.J.; Thomas, A.; Best, N.G.; Lunn, D. WinBUGS Version 1.4 Users Manual. MRC Biostatistics Unit. 2003. Available online: http://www.mrc-bsu.cam.ac.uk/bugs/ (accessed on 1 January 2020).

26. Brooks, S.P.; Gelman, A. General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* **1998**, *7*, 434.

27. Gelman, A.; Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*, 1st ed.; Cambridge University Press: New York, NY, USA; Cambridge, UK, 2006.

28. Bock, R.D.; Lieberman, M. Fitting a response model for n dichotomously scored items. *Psychometrika* **1970**, *35*, 179–197. [CrossRef]

29. Verhulst, P. Mathematical researches into the law of population growth increase. *Nouv. Mem. L'academie R. Sci.* **1845**, *18*, 1–45.

30. Kingsland, S. The Refractory Model: The Logistic Curve and the History of Population Ecology. *Q. Rev. Biol.* **1982**, *57*, 29–52. [CrossRef]

31. Marchetti, C. *Modeling Innovation Diffusion*; Henry, B., Ed.; Forecasting Technological Innovation; Kluwer Academic Publishing: Dordrecht, The Netherlands, 1991; pp. 55–77.

32. Liao, C.Y.; Podrázský, V.V.; Liu, G.B. Diameter and height growth analysis for individual White Pine trees in the area of Kostelec nad Černými lesy. *J. For. Sci.* **2003**, *49*, 544–551. [CrossRef]

33. Rosenblatt, F. *Principles of Neurodynamics*; Spartan: New York, NY, USA, 1962.

34. Schmidhuber, J. Deep Learning in Neural Networks: An Overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef] [PubMed]

35. Choi, J.; Harring, J.R.; Hancock, G.R. Latent Growth Modeling for Logistic Response Functions. *Multivar. Behav. Res.* **2009**, *44*, 620–645. [CrossRef]

36. Choi, J.; Chen, J.; Harring, J.R. Logistic Growth Modeling with Markov Chain Monte Carlo Estimation. *J. Mod. Appl. Stat. Methods* **2019**, *18*, 2–18. [CrossRef]

37. Hambleton, R.K.; Swaminathan, H. *Item Response Theory Principles and Applications*; Kluwer-Nijhoff Publishing: Boston, MA, USA, 1985.
38. van der Linden, W.J.; Hambleton, R.K. (Eds.) *Handbook of Modern Item Response Theory*; Springer: New York, NY, USA, 1997.
39. Mislevy, R.J. Recent developments in the factor analysis of categorical variables. *J. Educ. Stat.* **1986**, *11*, 3–31. [CrossRef]