Title
Poisson Diagnostic Classification Models: A Framework and an Exploratory Example

Journal
Educational and Psychological Measurement, 82(3)

ISSN
0013-1644

Authors
Liu, Ren
Liu, Haiyan
Shi, Dexin
et al.

Publication Date
2022-06-01

DOI
10.1177/00131644211017961

Peer reviewed

Article

Educational and Psychological Measurement
2022, Vol. 82(3) 506–516

DOI: 10.1177/00131644211017961
journals.sagepub.com/home/epm
SAGE

Poisson Diagnostic Classification Models: A Framework and an Exploratory Example

Ren Liu[1], Haiyan Liu[1], Dexin Shi[2] and Zhehan Jiang[3]

Abstract

Assessments with a large amount of small, similar, or often repetitive tasks are being used in educational, neurocognitive, and psychological contexts. For example, respondents are asked to recognize numbers or letters from a large pool of those and the number of correct answers is a count variable. In 1960, George Rasch developed the Rasch Poisson counts model (RPCM) to handle that type of assessment. This article extends the RPCM into the world of diagnostic classification models (DCMs) where a Poisson distribution is applied to traditional DCMs. A framework of Poisson DCMs is proposed and demonstrated through an operational dataset. This study aims to be exploratory with recommendations for future research given in the end.

Keywords

diagnostic classification model, Poisson distribution, count data

Diagnostic classification models (DCMs) classify examinees into groups of possessing or not possessing a set of attributes based on their item responses. New DCMs are burgeoning every day to handle complex and/or unique needs of assessments in

[1]University of California, Merced, CA, USA
[2]University of South Carolina, Columbia, CA, USA
[3]Peking University, Beijing, China

Corresponding Authors:
Ren Liu, Quantitative Methods, Measurement, and Statistics (QMMS), University of California, Merced, 5200 N Lake Rd, Merced, CA 95343, USA.
Email: rliu45@ucmerced.edu

Zhehan Jiang, Institute of Medical Education & National Center for Health Professions Education Development, Peking University, Beijing 100191, China.
Email: jiangzhehan@gmail.com

educational and behavioral sciences. In this article, we consider situations where test scores are count variables (e.g., Doebler et al., 2014; Jansen, 1997; Spray, 1990). For example, examinees are asked to circle correctly spelled words in a 200-word test, and their scores are the number of words they correctly pick out. Another example is that examinees read aloud a passage and the number of errors is counted (Rasch, 1960). It could also be that examinees are asked to check yes/no from 50 simple conditions regarding stress and anxiety such as blush, nightmare, diarrhea, and sweat, and their scores are the number of conditions that they select (Taylor, 1953). A shared feature of the aforementioned situations is that the number of items is large, and items are either the same or very similar.

Current DCMs are item-based models where the unit of analysis is each examinee's response on each item, just like most item response theory (IRT) models. As a result, each item is associated with at least two parameters under a DCM. This happens when the item only measures one attribute, is modeled under the simplest DCM, and is scored in a binary fashion. When an item measures more attributes, is modeled under a more complex DCM, or scored in a polytomous fashion, the item could be associated with 10, 20, or even more parameters. Using existing DCMs to model the example situations above brings in an enormously large amount of and unnecessary parameters. Under those situations, it may be more appropriate to model the counts of total successes rather than the responses on each individual item. Counts are often modeled as coming from a Poisson distribution. The purpose of this article is to develop a family of Poisson DCMs (PDCMs) where counts of successes or errors on a set of tasks is modeled. The family of PDCMs closely mirrors the Rasch Poisson counts model (RPCM; Rasch, 1960) in IRT and extends its basic structure into the DCMs. Before introducing the PDCMs, we first review the basics of the Poisson distribution and the RPCM in the next section.

Model Development

Poisson Distribution

In 1830, Siméon-Denis Poisson, a French mathematician, developed a function to describe the number of times an event could happen in a fixed time period when you know how often the event has happened. That function is known as the Poisson distribution. The Poisson distribution defines the probability that the event happens s times as

$$P(s) = \frac{\lambda^s exp^{-\lambda}}{s!}, \tag{1}$$

where λ is the expected number of times that an event would happen in the fixed time period. The λ parameter, known as the rate parameter in a Poisson distribution, is often a multiplicative function of both the number of tries (n) and the probability of events happening (p): $\lambda = np$. The Poisson distribution assumes that each event is independent and distributed with a mean and variance of λ.

Rasch Poisson Counts Model

In 1960, George Rasch applied the Poisson distribution to model the number of errors students make during oral reading tasks and thus developed the RPCM. Although it is one of the earliest models that Rasch has developed, the RPCM has received less attention than other binary or polytomous Rasch models. Suppose we ask students to read 200 words in 10 blocks where each block has 20 words, and we record the number of successes in each block. In the RPCM, the unit of analysis is no longer students' correct/incorrect responses on each word; instead, an item i is defined as a block, and a student's score s on item i is the total number of successes in that block. In other words, we could think of an item in the RPCM as a subtest or a collection of tasks. The RPCM defines the probability of obtaining a score of s on item i given a unidimensional latent trait θ for examinee e as

$$P(X_{ei} = s|\theta_e) = \frac{\lambda_{ei}^{s} exp^{-\lambda_{ei}}}{s!}, \tag{2}$$

where $\lambda_{ei} = \theta_e b_i$, and b_i represents the easiness of an item (i.e., easier items are associated with larger b_i values). Comparing Equations 2 to 1, we should be able to see that the RPCM is a direct multivariate application of the Poisson distribution. One problem with the current λ_{ei} construction is that the person parameter (θ_e) and item parameter (b_i) are tied together using the multiplication. We are often interested in obtaining separate person and item parameter estimates, and most IRT models estimate those two types of parameters separately through an addition such as: $\theta_e + b_i$. In addition, the λ_{ei} represents the expected score that must be nonnegative, but both θ_e and b_i do not have boundaries. To solve those two problems, Haberman (1978) used a natural log as a link with an additive specification to replace the multiplicative specification such that:

$$\lambda_{ei} = \exp\left(\tilde{\lambda}_{ei}\right) = \exp[\ln(\theta_e b_i)] = \exp[\ln(\theta_e) + \ln(b_i)] = \exp\left(\tilde{\theta}_e + \tilde{b}_i\right). \tag{3}$$

We can obtain the original θ_e and b_i parameters by applying the exponential function such that $\theta_e = \exp(\tilde{\theta}_e)$ and $b_i = \exp\left(\tilde{b}_i\right)$.

The Poisson Diagnostic Classification Model Framework

In essence, DCMs are confirmatory latent class models with different parameterizations of the measurement component. The basic structure a DCM can be written as

$$P(\boldsymbol{X}_e = \boldsymbol{x}_e) = \sum_{c=1}^{C} \xi_c \prod_{i=1}^{I} \pi_{i,e}, \tag{4}$$

where $c = 1, \ldots, C$ indexes predefined latent classes, ξ_c represents the proportion of examinees in class c, and $\pi_{i,e}$ denotes the probability of examinee e's response on item i. For example, existing DCMs for binary items parameterize $\pi_{i,e}$ into

$P(X_i = 1|\boldsymbol{\alpha}_c)$ based on different theories. To develop the family of PDCMs to account for count data, we parameterize $\pi_{i,e}$ using the Poisson distribution. In this article, we use $k = 1, \ldots, K$ to index binary latent traits (aka attributes) and $\boldsymbol{\alpha}_c = \{\alpha_1, \ldots, \alpha_K\}$ to index attribute profiles for latent class c. The information about item–attribute relationship is contained in an item-by-attribute incidence matrix, commonly referred to as the Q-matrix (Tatsuoka, 1983), where an entry $q_{ik} = 1$ when item i measures attribute k, and $q_{ik} = 0$ otherwise. The general form of the PDCM defines the probability of examinees in attribute profile c obtaining a score of s on item i as

$$P(X_i = s|\boldsymbol{\alpha}_c) = \frac{\lambda_{ci}^s exp^{-\lambda_{ci}}}{s!}, \tag{5}$$

$$\lambda_{ci} = \boldsymbol{\omega}_i^T \mathbf{h}(\boldsymbol{\alpha}_c, \mathbf{q}_i)\omega_{0,i}, \tag{6}$$

$$\boldsymbol{\omega}_i^T \mathbf{h}(\boldsymbol{\alpha}_c, \mathbf{q}_i) = \sum_{k=1}^{K} \omega_{1,i,k}(\alpha_{c,k} q_{i,k}) + \sum_{k=1}^{K-1} \sum_{k'=k+1}^{K} \omega_{2,i,k,k'}(\alpha_{c,k}\alpha_{c,k'} q_{i,k} q_{i,k'}) + \cdots \tag{7}$$

$$\begin{aligned}\lambda_{ci} &= \exp\left[\ln\left(\boldsymbol{\omega}_i^T \mathbf{h}(\boldsymbol{\alpha}_c, \mathbf{q}_i)\omega_{0,i}\right)\right] = \exp\left[\ln\left(\boldsymbol{\omega}_i^T \mathbf{h}(\boldsymbol{\alpha}_c, \mathbf{q}_i)\right) + \ln(\omega_{0,i})\right] \\ &= \exp\left[\widetilde{\boldsymbol{\omega}_i^T}(\boldsymbol{\alpha}_c, \mathbf{q}_i) + \tilde{\omega}_{0,i}\right].\end{aligned} \tag{8}$$

Let us explain each equation. Equation 5 is the item response function, which takes on a form similar to the RPCM in Equation 2. Equation 6 describes the parameterization of the λ_{ci}, where we use $\omega_{0,i}$ to denote the baseline easiness of item i for examinees that do not master the attribute(s) item i measures, and $\boldsymbol{\omega}_i^T \mathbf{h}(\boldsymbol{\alpha}_c, \mathbf{q}_i)$ is defined in Equation 7 as all the possible main and interaction effects that examinees in $\boldsymbol{\alpha}_c$ could impose on item i. The definition of λ_{ci} is obtained by replacing θ_e in the RPCM by $\boldsymbol{\omega}_i^T \mathbf{h}(\boldsymbol{\alpha}_c, \mathbf{q}_i)$, and b_i by $\omega_{0,i}$. Similar to the approach that the multiplication is dealt with in the RPCM, we use the additive log-linear function to separate the two components of the λ_{ci} as shown in Equation 8. We can obtain the original form of parameters by $\exp\left[\widetilde{\boldsymbol{\omega}_i^T}(\boldsymbol{\alpha}_c, \mathbf{q}_i)\right] = \omega_i^T(\boldsymbol{\alpha}_c, \mathbf{q}_i)$, and $\exp(\tilde{\omega}_{0,i}) = \omega_{0,i}$.

This additive specification in Equation 8 allows us to connect the PDCM to other DCMs that are based on generalized linear mixed models such as the log-linear cognitive diagnosis model (LCDM; Henson et al., 2009). The LCDM is the most general DCM for binary items that subsumes most earlier DCMs such as the ''deterministic inputs, noisy, and gate'' (DINA) model (Haertel, 1989; Junker & Sijtsma, 2001) and the linear logistic model (LiLM; Maris, 1999). The LCDM defines the probability of examinees in attribute profile c correctly answering item i as

$$P(X_i = 1|\boldsymbol{\alpha}_c) = \frac{\exp\left[\omega_{0,i} + \boldsymbol{\omega}_i^T \mathbf{h}(\boldsymbol{\alpha}_c, \mathbf{q}_i)\right]}{1 + \exp\left[\omega_{0,i} + \boldsymbol{\omega}_i^T \mathbf{h}(\boldsymbol{\alpha}_c, \mathbf{q}_i)\right]}. \tag{9}$$

Comparing Equations 9 to 8, we should be able to see the same set up of the $\omega_{0,i} + \boldsymbol{\omega}_i^T \mathbf{h}(\boldsymbol{\alpha}_c, \mathbf{q}_i)$ between the LCDM and the PDCM.

To estimate parameters in the PDCM, we impose two types of constraint, one for identifiability where some estimation approaches need and the other one for combating label switching. For identifiability, we could adapt the approach developed in Jansen and van Duijn (1992) and impose

$$\sum_{i}^{I} \omega_{0,i} = 1,$$

$$\sum_{i}^{I} \omega_{1,i,k} = 1 \forall k,$$

$$\sum_{i}^{I} \omega_{2,i,k,k'} = 1 \forall k, k',$$

and for all higher order interactions. Alternatively, one could put constraints on the parameter distributions, which will be demonstrated in the operational study. To combat label switching, we force $\omega_{1,i,k} \geq 0 \forall i, k$ so that mastering more attributes does not decrease the mean of the Poisson distribution (i.e., λ_{ci}). In other words, this constraint ensures that the average number of successes should increase or at least remain the same if an examinee masters more attributes.

Models Under the PDCM Framework

The PDCM provides a modeling framework that could accommodate item response functions that reflect specific theories. The family of PDCMs share the same structure in Equation 5 but with different parameterizations of the λ_{ci} parameter. For example, the PDCM-DINA model is based on the DINA model and reflects the theory that not mastering an attribute cannot be compensated for by mastering another attribute regarding the total score. In the PDCM-DINA model,

$$\lambda_{ci} = \left(\omega_{1,i} \prod_{k=1}^{K} \alpha_{c,k}^{q_{i,k}} \right) \omega_{0,i}, \tag{10}$$

$$\lambda_{ci} = \exp\left\{ \ln \left[\left(\omega_{1,i} \prod_{k=1}^{K} \alpha_{c,k}^{q_{i,k}} \right) \omega_{0,i} \right] \right\} = \exp\left(\tilde{\omega}_{1,i} \prod_{k=1}^{K} \alpha_{c,k}^{q_{i,k}} + \tilde{\omega}_{0,i} \right). \tag{11}$$

We can see that the λ_{ci} in the PDCM-DINA contains only two parameters per item $\omega_{1,i}$ and $\omega_{0,i}$ regardless of how many attributes an item measures. This could help reduce estimation burdens comparing to the general PDCM when an item measures many attributes. Similarly, the PDCM-LLM, based on the LLM, reduces estimation burdens through removing all the interaction effects between attributes. In the PDCM-LLM,

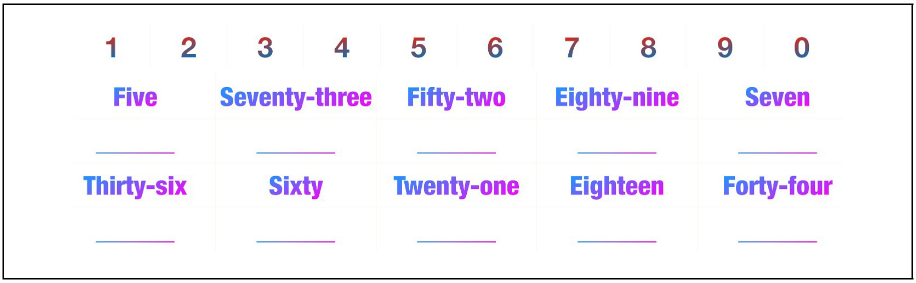

Figure 1. An example block in the operational study.

$$\lambda_{ci} = \left[\sum_{k=1}^{K} \omega_{1,i,k}(\alpha_{c,k} q_{i,k})\right] \omega_{0,i}, \tag{12}$$

$$\lambda_{ci} = \exp\left\{\ln\left[\left[\sum_{k=1}^{K} \omega_{1,i,k}(\alpha_{c,k} q_{i,k})\right] \omega_{0,i}\right]\right\} = \exp\left(\sum_{k=1}^{K} \tilde{\omega}_{1,i,k}(\alpha_{c,k} q_{i,k}) + \tilde{\omega}_{0,i}\right). \tag{13}$$

We can see that the λ_{ci} contains $k_i + 1$ parameters (i.e., k_i $\omega_{1,i,k}$ parameters and one $\omega_{0,i}$) for item i. In addition to the two examples here, most traditional DCMs could be applied to the PDCM with λ_{ci} specifications. In applications, both theory and model fitting results could provide information the selection of models in the PDCM family.

Operational Study

To demonstrate how to use the PDCM in operational settings, we obtained data from 808 kindergarten students in China on an English Recognition Assessment. The assessment measures their abilities to recognize basic English in three domains: numbers, colors, and objects. Each domain has eight blocks, each containing 10 English words. Students need to drag the numbers/colors/objects into the blanks that match the English words shown on the screen. An example block is given in Figure 1. In this example, we can see that it may not be adequate to model each word as an item because they are so simple and similar. Instead, we could count the total number of correct or incorrect answers in each block (which is treated as an item) and fit the proposed PDCM. Figure 2 shows the frequencies of examinees' incorrect counts in each block under each domain. Because most examinees have got most English words correct in each block, we chose to count the number of incorrect answers in each block as the input for the s parameter in Equation 5. When interpreting the results, we flipped the binary definitions of each attribute where $\alpha_k = 0$ represents mastery of α_k and $\alpha_k = 1$ represents nonmastery of α_k.

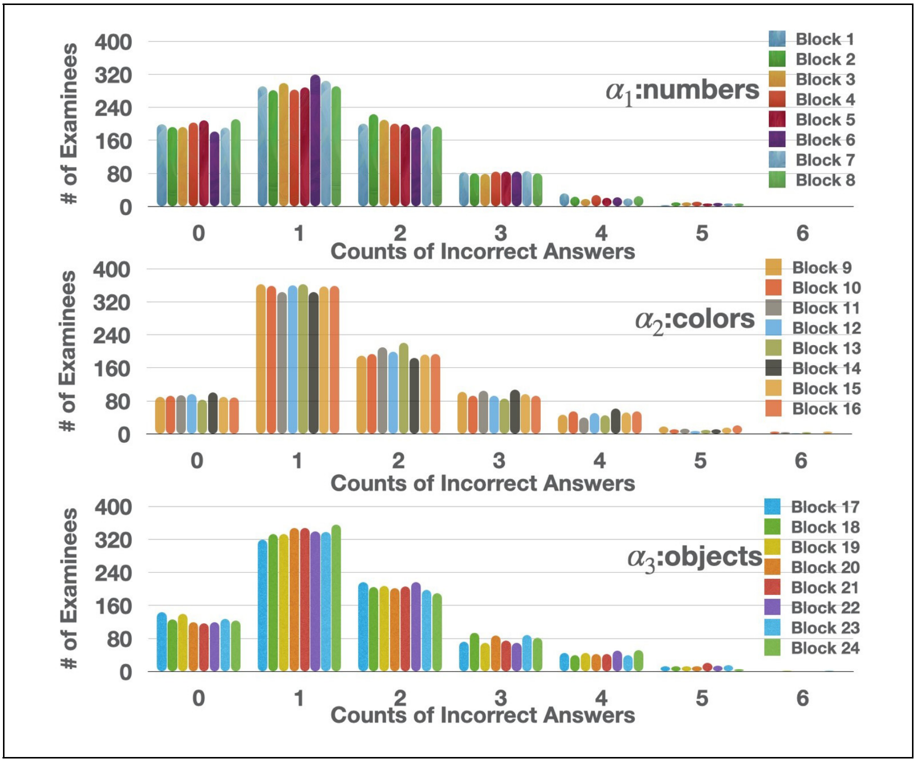

Figure 2. Distributions of the counts of examinees' incorrect answers in each block.

We estimated the PDCM parameters using Hamiltonian Monte Carlo (HMC) algorithms in Stan (Carpenter et al., 2017). The Stan syntax can be found in the Supplemental Material. Specifically, we estimated a vector of class membership probabilities: ξ_c, 24 transformed intercept parameters: $\tilde{\omega}_{0,i}$, and 24 transformed main effect parameters: $\tilde{\omega}_{1,i}$. The prior distributions, similar to those used in de la Torre and Douglas (2004) and Liu and Liu (2020), were specified as

$$\xi_c \sim \text{Dirichlet}(2),$$

$$\tilde{\omega}_{0,i} \sim \text{Normal}(0, 2),$$

$$\tilde{\omega}_{1,i} \sim \text{Normal}(0, 2).$$

Different priors may be selected based on theories and properties of the dataset, which may influence parameter estimates. Four Markov chains were specified with each a length of 20,000, where the first 10,000 draws were discarded and the rest 10,000 were used for inference. Gelman-Rubin's $\hat{R}$ values (Gelman & Rubin, 1992)

Table 1. Mean and Standard Deviation of the Posterior Distribution for Each Item Parameter in the Operational Study.

	Item	Mean		Standard deviation		Transformed λ parameters	
		$\tilde{\omega}_{0,i}$	$\tilde{\omega}_{1,i}$	$\tilde{\omega}_{0,i}$	$\tilde{\omega}_{1,i}$	$\lambda_{\alpha=0,i}$	$\lambda_{\alpha=1,i}$
α_1: numbers	1	−0.987	0.674	−0.987	0.674	0.373	0.731
	2	−0.930	0.689	−0.930	0.689	0.395	0.786
	3	−0.929	0.658	−0.929	0.658	0.395	0.763
	4	−1.012	0.693	−1.012	0.693	0.364	0.727
	5	−1.056	0.660	−1.056	0.660	0.348	0.673
	6	−0.845	0.659	−0.845	0.659	0.429	0.830
	7	−0.912	0.659	−0.912	0.659	0.402	0.776
	8	−1.082	0.655	−1.082	0.655	0.339	0.652
α_2: colors	9	−0.333	0.801	−0.333	0.801	0.717	1.598
	10	−0.341	0.812	−0.341	0.812	0.711	1.602
	11	−0.351	0.805	−0.351	0.805	0.704	1.575
	12	−0.364	0.773	−0.364	0.773	0.695	1.506
	13	−0.302	0.780	−0.302	0.780	0.739	1.613
	14	−0.382	0.823	−0.382	0.823	0.682	1.554
	15	−0.331	0.825	−0.331	0.825	0.718	1.639
	16	−0.323	0.819	−0.323	0.819	0.724	1.642
α_3: objects	17	−0.598	0.745	−0.598	0.745	0.550	1.158
	18	−0.500	0.757	−0.500	0.757	0.606	1.293
	19	−0.576	0.741	−0.576	0.741	0.562	1.178
	20	−0.465	0.749	−0.465	0.749	0.628	1.329
	21	−0.454	0.765	−0.454	0.765	0.635	1.365
	22	−0.469	0.757	−0.469	0.757	0.626	1.334
	23	−0.516	0.763	−0.516	0.763	0.597	1.280
	24	−0.496	0.730	−0.496	0.730	0.609	1.263

for each parameter were smaller than 1.00, suggesting convergence. Posterior predictive *p*-values (PPPs; Gelman et al., 2013) was 0.55, suggesting good model-data fit.

Table 1 lists the mean and standard deviation of the posterior distribution for each item parameter. We also computed the λ_{ci} in Equation 8 using the mean values of $\tilde{\omega}_{0,i}$ and $\tilde{\omega}_{1,i}$. Using the estimated λ_{ci}, we could compute the probability of getting different incorrect counts in each block. Take Block 1 (i.e., $i=1$) as an example; Block 1 is measuring α_1: "numbers," and we could compute the probability of getting 0 incorrect answer, 1 incorrect answer, and 2 incorrect answers using Equation 5, $P(X_i=s|\boldsymbol{\alpha}_c)=\frac{\lambda_{ci}^s exp^{-\lambda_{ci}}}{s!}$, that is,

$P(X_{i=1}=0|\alpha_1=0)=\frac{0.373^0 exp^{-0.373}}{0!}=\frac{exp^{-0.373}}{1}=0.689$ for masters,

$P(X_{i=1}=0|\alpha_1=1)=\frac{0.731^0 exp^{-0.731}}{0!}=\frac{exp^{-0.731}}{1}=0.481$ for non-masters,

$P(X_{i=1}=1|\alpha_1=0)=\frac{0.373^1 exp^{-0.373}}{1!}=\frac{0.373 exp^{-0.373}}{1}=0.257$ for masters,

$P(X_{i=1}=1|\alpha_1=1)=\frac{0.731^1 exp^{-0.731}}{1!}=\frac{0.731 exp^{-0.731}}{1}=0.352$ for non-masters,

$P(X_{i=1}=2|\alpha_1=0)=\frac{0.373^2 exp^{-0.373}}{2!}=\frac{0.139 exp^{-0.373}}{1}=0.010$ for masters,

Figure 3. Model predicted probabilities of the counts of examinees' incorrect answers in each block.

and

$P(X_{i=1} = 2|\alpha_1 = 1) = \frac{0.731^2 exp^{-0.731}}{2!} = \frac{0.534 exp^{-0.731}}{1} = 0.257$ for non-masters.

Similarly, we computed those values for each block and displayed the results in Figure 3. Figure 3 only showed the probabilities of up to six incorrect answers because the probabilities for more than six incorrect answers became too small. However, one could compute the probabilities of getting any numbers of incorrect answers up to the total number of blocks. The overall trend in Figure 3 is that for masters, the probability of getting more incorrect answers was constantly decreasing for items that measure any of the three attributes. For non-masters, the probability of getting more incorrect answers was decreasing for items that measure α_1: ''numbers,'' but the peak for α_2: ''colors'' and α_3: ''objects'' was at one incorrect answer. Comparing across the three attributes, the probabilities of getting incorrect answers tell us that α_1 was easier than α_2, which was easier than α_3.

Discussion

This article provided a framework for modeling count data under the DCM framework and demonstrated its uses and interpretations with an operational dataset. The PDCM framework can be especially useful for assessments with many small and repetitive units. For example, in education, teachers give out words or numbers to test whether students have mastered them; in psychology, simple recognition tasks are used to test respondents' attention, memory, or thinking skills. Integrating the Poisson distribution into the DCMs offers tools for teachers to decide student mastery on specific knowledge and for psychologists to classify respondents with mental disorders.

The current study aims to be an exploratory piece of work that taps into DCM with count data. Future work is needed on at least the following aspects. First, although it is expected that the PDCM holds similar characteristics as other psychometric models where larger sample sizes and/or longer test length and Q-matrices with fewer cross loadings produce more accurate results, simulation studies could be conducted to verify that. Second, some operational assessments of this type may have time limits on each task. Extending the current PDCM to include time factor could be studied. Specifically, one could consider $\lambda_{ci} = \tau_i \boldsymbol{\omega}_i^T \mathbf{h}(\boldsymbol{\alpha}_c, \mathbf{q}_i)\omega_{0,i}$, where τ_i represents the time limit for item i. Third, beyond the current HMC estimation method, other estimation methods such as the maximum likelihood estimation could be studied. Fourth, operational psychometric topics such as differential item functioning could be studied with respect to the PDCM. Ultimately, much more work is needed as the PDCM becomes more useful with respect to its specific type of assessment.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Supplemental Material

Supplemental material for this article is available online.

ORCID iDs

Ren Liu https://orcid.org/0000-0002-6708-4996
Dexin Shi https://orcid.org/0000-0002-4120-6756

References

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*(3), 333-353.

Doebler, A., Doebler, P., & Holling, H. (2014). A latent ability model for count data and application to processing speed. *Applied Psychological Measurement*, *38*(8), 587-598. https://doi.org/10.1177/0146621614543513

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Chapman & Hall/CRC Press. https://doi.org/10.1201/b16018

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457-511. https://doi.org/10.1214/ss/1177011136

Haberman, S. J. (1978). *Analysis of qualitative data*. Academic Press.

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, *26*(4), 301-321. https://doi.org/10.1111/j.1745-3984.1989.tb00336.x

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*(2), 191-210. https://doi.org/10.1007/s11336-008-9089-5

Jansen, M. G. H. (1997). Rasch's model for reading speed with manifest exploratory variables. *Psychometrika*, *62*, 393-409. https://doi.org/10.1007/BF02294558

Jansen, M. G. H., & van Duijn, M. A. (1992). Extensions of Rasch's multiplicative Poisson model. *Psychometrika*, *57*, 405-414. https://doi.org/10.1007/BF02295428

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*(3), 258-272. https://doi.org/10.1177/01466210122032064

Liu, R., & Liu, H. (2020). Nested diagnostic classification models for multiple-choice items. *British Journal of Mathematical and Statistical Psychology*, *74*(2), 257-285. https://doi.org/10.1111/bmsp.12214

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, *64*, 187-212. https://doi.org/10.1007/BF02294535

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. University of Chicago Press.

Spray, J. A. (1990). One-parameter item response theory models for psychomotor tests involving repeated independent attempts. *Research Quarterly for Exercise and Sport*, *61*(2), 162-168. https://doi.org/10.1080/02701367.1990.10608670

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20*(4), 345-354.

Taylor, J. (1953). A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology*, *48*(2), 285-290. https://doi.org/10.1037/h0056264