Computational Statistics and Data Analysis 68 (2013) 361–374

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Multivariate zero-inflated modeling with latent predictors: Modeling feedback behavior

Jean-Paul Fox *

University of Twente, Faculty of Behavioral Sciences, P.O. Box 217, 7500 AE Enschede, The Netherlands

HIGHLIGHTS

- We propose a multivariate zero-inflated Poisson–Gamma model for counts and times.
- We studied feedback behavior in a computer-based assessment.
- High ability students were more likely to consult feedback.
- The number of feedback pages consulted was negatively related to achievement.
- Fast working students were not likely to consult feedback.

ARTICLE INFO

Article history:
Received 28 November 2012
Received in revised form 1 July 2013
Accepted 1 July 2013
Available online 11 July 2013

Keywords:
Feedback behavior
Latent predictors
Multivariate zero-inflated count data
Item response times
Poisson-log normal model
Poisson–Gamma model

ABSTRACT

In educational studies, the use of computer-based assessments leads to the collection of multiple outcomes to assess student performance. The student-specific outcomes are correlated and often measured in different scales, such as continuous and count outcomes. A multivariate zero-inflated model with random effects is proposed and adapted for the challenging situation where the multiple outcomes are zero-inflated and possibly right truncated. The joint model consists of a Bernoulli component to deal with the problem of extra zeros, and a multivariate truncated component to model correlated mixed response outcomes from the same subject. In a Bayesian modeling approach, MCMC methods are used for parameter estimation. Using a simulation study, it is shown that the within-individual correlation between counts can be accurately estimated together with the other model parameters. The multivariate zero-inflated model is applied to a computer-based feedback study about computer literacy, where first-year bachelor students were given the opportunity to receive additional feedback. The total number of feedback pages visited and the total feedback processing time are modeled using a Poisson and a Gamma distribution, respectively. The joint modeling framework is extended to incorporate explanatory latent variables (student performance and speed of working), to explore individual heterogeneity in feedback behavior in a computer-based assessment.

1. Introduction

Computer-based assessment (a.k.a. computer-administered testing), where students are requested to answer items in a computer environment, is receiving increasing attention. Various advantages of computer-based assessment have been exploited such as improved reliability, improved question styles using interactive multimedia technology, and testing on demand, among other things (e.g., van der Linden and Glas, 2010). Recently, the opportunity to provide instant feedback to students has been further developed to improve learning. In different computer-based assessment studies, feedback

* Tel.: +31 53 4893326.
E-mail address: J.P.Fox@utwente.nl.

0167-9473/$ – see front matter
http://dx.doi.org/10.1016/j.csda.2013.07.003

information was integrated to inform students immediately about their deficiencies given the test results. The main object was to investigate the effects of feedback on students' learning outcomes (e.g., Hattie and Timperly, 2007; van der Kleij et al., 2012).

The present study concerns a computer-based formative assessment (CBFA), where first-year bachelor students were given the opportunity to consult feedback after completing an information literacy test. Interest was focused on student-specific propensity to consult feedback and student-specific characteristics of feedback behavior such as the expected number of pages visited and the expected total time to process the feedback information, also referred to as the total reading time. The assessment data provide information about student achievement and working speed.

From a statistical point of view, several important features of the feedback data need to be considered. First, the data are potentially zero-inflated, where a substantial group of students, around 42%, did not consult feedback for different reasons (e.g., time limitations, motivation, not familiar with computer-based assessment). Second, the restricted number of test items leads to right truncated feedback-use counts. Third, the probability of feedback use is most likely correlated with the expected number of consulted feedback pages and the expected time reading feedback. This correlation should be taken into account to avoid biased parameter estimates. Fourth, within-subject correlation needs to be addressed, since the number of pages visited and the total time reading are clustered by subject. Fifth, the total number of feedback pages consulted (count data) and the total feedback reading time (positive continuous data) are measured on different scales. Finally, individual latent predictors, achievement and speed of working, are measured using a response time item response model (RTIRT; Klein Entink et al., 2009; van der Linden, 2007), which requires a proper treatment of the statistical error in the measurements when using them as explanatory variables.

Here, a multivariate zero-inflated model with subject-specific random effects is presented that addresses the specific features of the data. The real-data study about feedback use will be used throughout the paper to illustrate the various modeling issues. The model consists of two components, where one component addresses the excess zeros and the other multivariate component describes the variability of the counts (total visited pages) and the continuous times (total feedback reading time). A truncated Poisson distribution is used to model the counts and a Gamma distribution is used to model the times, where subject-specific random effects are used to model the unobserved between-subject variation. The log of the mean parameters, representing the expected total feedback use and expected total processing time, are considered to be random effects and assumed to be multivariate normally distributed to capture the within-subject correlation. This joint model is referred to as a multivariate zero-inflated Poisson–Gamma model, where the random effect mean parameters are referred to as subject-specific rates.

Several motivations can be given for the joint modeling framework. Inferences can be made for the original mixed outcomes, the joint model implies separate univariate models, marginal inferences can be directly made, and the random effect structure avoids heavy computational problems since it reduces the issue of dimensionality.

In the field of health research, where it is more common to measure mixed outcomes, the joint modeling of mixed outcomes received considerable attention. Many applications consider the modeling of multivariate longitudinal outcomes and a time-to-event outcome (e.g., McCulloch, 2008; Rizopoulos et al., 2010; Yang et al., 2007). Furthermore, in practice, observations might be zero-inflated or censored such that values below or above a limit are not observed (Mihaylova et al., 2011; Rose et al., 2006). For example, in HIV/AIDS studies viral loads below a certain limit cannot be measured and are subject to left censoring (Hu et al., 2011). In the present modeling framework, of typical interest would be, in the case of informative dropouts, the joint modeling of the censored viral load measurements and time-to-event data. When considering healthcare cost data (as the aggregated total cost of different sources) and the number of interventions (i.e. number of times a medical intervention was required during a time period) a joint modeling approach will account for within-subject correlated observations. A zero-inflated multivariate modeling approach was required when interventions were sparse for subjects in the population. In other research fields, zero-inflated multivariate (count) data also arise naturally in different applications. For example, in economics, purchases over time of different households and products are likely to be correlated within households. And purchase observations of products that are rarely used are zero-inflated.

The present approach builds on work on zero-inflated (count) models. This includes the zero-inflated Poisson model, Lambert (1992) and Loeys et al. (2012) for a more general introduction, where recently Bayesian alternatives have been proposed (see Congdon, 2005; Gelman et al., 2004). Wang (2010) also proposed a model to handle zero-inflated count data. Zero-inflated Poisson models with random effects were also considered by Min and Agresti (2005) and Rabe-Hesketh and Skrondal (2007). The excess zeros will be explicitly modeled through a Bernoulli model, where the success probability is defined as feedback use. This will make it possible to investigate the heterogeneity in the feedback use and student characteristics can be used to explain individual differences in the propensity to use feedback.

The multivariate zero-inflated model is further extended by introducing latent student predictors (achievement and speed of working) as covariates. Latent student characteristics are measured using a response time item response model (RTIRT; Klein Entink et al., 2009; van der Linden, 2007) and used as explanatory variables (e.g., Fox and Glas, 2003; Skrondal and Rabe-Hesketh, 2004). A Bayesian estimation approach is adopted, which supports joint estimation of all model parameters using MCMC simulation techniques.

The remainder of this paper is organized as follows. In the next section, the proposed multivariate zero-inflated mixture model is described using the real-data study as an example throughout the paper. Then, the measurement model for the latent predictors is described. An MCMC algorithm is proposed, which represents the sampling steps related to the different model components. Then, the model is applied to the Dutch feedback data. Finally, the last section provides a discussion and outlines directions for further research.

2. Multivariate zero-inflated random effect models

A formal description of the modeling framework is given on the basis of the feedback study. A total of n students are considered that have an opportunity to consult feedback for each item $k = 1, \ldots, K$. Let Y_i^f denote the total number of feedback pages visited ($k = 0, 1, 2, \ldots, K$) by student i ($i = 1, \ldots, n$). The observed number of feedback pages of student i will be assumed to be Poisson distributed with feedback rate $\lambda_i^{(f)}$.

The observed discrete count data may contain excess zeros relative to what can be expected according to the Poisson distribution. This may occur when a significant group of students do not open any of the feedback pages. This group is not interested in feedback and will represent a zero state. Other students may ignore the feedback this time, but they might consult feedback on another test occasion. Following Lambert (1992), the relevant distribution is a mixture of the Poisson distribution and a degenerate distribution at zero. The zero-inflated Poisson model for the total number of feedback pages opened of subject i is represented by

$$Y_i^f \sim \begin{cases} 0, & \text{with probability } 1 - \phi_i \\ Poisson\left(\lambda_i^{(f)}\right), & \text{with probability } \phi_i, \end{cases} \tag{1}$$

where $1 - \phi_i$ is the probability of an excess zero of student i. It follows that the probability of a zero response and a non-zero response is given by, respectively,

$$P\left(Y_i^f = 0 \mid \lambda_i = \lambda_i^{(f)}, Y_i^f \leq K\right) = (1 - \phi_i) + \phi_i e^{-\lambda_i}/C_i(K), \tag{2}$$

$$P\left(Y_i^f = j \mid \lambda_i = \lambda_i^{(f)}, Y_i^f \leq K\right) = \phi_i \frac{e^{-\lambda_i}\lambda_i^j}{j!} \Big/ C_i(K), \tag{3}$$

where $j = 1, 2, \ldots, K$. Here,

$$C_i(K) = P\left(Y_i^f \leq K \mid \lambda_i = \lambda_i^{(f)}\right) = \sum_{j=0}^{K} \frac{e^{-\lambda_i}\lambda_i^j}{j!}$$

is the truncation probability to account for the upper bound of K items.

Let T_i^f denote student i's total time (in seconds) reading or processing the feedback information. This feedback time equals zero when student i does not open any feedback page. Subjects not opening feedback pages generate a substantial proportion of zeros. Therefore, feedback time data are modeled as a mixture of a degenerate distribution at zero (with success probability $1 - \phi_i$) and a distribution for the non-zero observations.

The zero-inflated time observations are assumed to be zero-inflated Gamma distributed, which is represented by

$$P\left(T_i^f = 0 \mid \lambda_i = \lambda_i^{(t)}\right) = (1 - \phi_i), \tag{4}$$

$$P\left(T_i^f = t_i \mid \lambda_i = \lambda_i^{(t)}, \nu\right) = \frac{\phi_i}{\left(\frac{2\lambda_i}{\nu_i}\right)^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)} t_i^{(\nu-2)/2} \exp\left(\frac{-t_i\nu}{2\lambda_i}\right), \tag{5}$$

where $0 < t_i < \infty$ and $\Gamma(\cdot)$ denotes the Gamma function. Note that in this parametrization of the Gamma distribution the shape parameter is $\nu/2$ and the scale parameter $2\lambda_i^{(t)}/\nu$. Then, the expected feedback time is equal to $\lambda_i^{(t)}$ and the variance to $2\lambda_i^{2(t)}/\nu$.

When a test is administered with a time limit, the total time is restricted to be lower than the time limit. In that case, the total time is truncated Gamma distributed, where the truncation probability accounts for the time limit.

In some cases, the times are coarsely observed, when, for example, discrete times are recorded as proxies of the continuous observations. In the present study, relatively small total feedback times were stored in whole seconds. These discrete time observations can also be modeled by a Poisson distribution, where a Poisson rate parameter $\lambda_i^{(t)}$ can represent the expected total time (in seconds) of subject i's feedback time. In the same way as Eq. (1), the zero-inflated Poisson model for the feedback (reading) time is given by

$$P\left(T_i^f = 0 \mid \lambda_i = \lambda_i^{(t)}\right) = (1 - \phi_i) + \phi_i e^{-\lambda_i}, \tag{6}$$

$$P\left(T_i^f = j \mid \lambda_i = \lambda_i^{(t)}\right) = \phi_i \frac{e^{-\lambda_i}\lambda_i^j}{j!}, \quad j = 1, 2, \ldots \tag{7}$$

In the Gamma distribution and in the Poisson distribution, (5) and (7), the parameter $\lambda_i^{(t)}$ represents the expected total reading time of subject *i*. The subject parameter $\lambda_i^{(f)}$ represents the expected number of feedback pages that are visited. Both subject parameters $\lambda_i^{(f)}$ and $\lambda_i^{(t)}$ are restricted to be positive and assumed to be independently distributed over subjects, but it is to be expected that they are nested within the subject. For example, the total amount of feedback time increases when opening more feedback pages.

The subject parameters are referred to as subject's rate parameters and the within-subject correlation between the rates are modeled with a multivariate log-normal distribution. That is, the logarithm of the subject's rates are multivariate normally distributed,

$$\left(\log(\lambda_i^{(f)}), \log(\lambda_i^{(t)})\right)^t \mid \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{N}_2\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right). \quad (8)$$

The student-specific probability of not opening any feedback page equals $1 - \phi_i$, which can be explicitly modeled to explore and explain differences between subjects in their use of feedback information. Let G_i denote the class-membership variable, where $G_i = 0$ when an excess zero is given and $G_i = 1$ otherwise. The probability ϕ_i represents the probability of a non-zero observation using a logistic probability model,

$$\phi_i = P\left(G_i = 1 \mid \boldsymbol{\alpha}\right) = \frac{\exp(\mathbf{x}_i^t\boldsymbol{\alpha})}{1 + \exp(\mathbf{x}_i^t\boldsymbol{\alpha})}, \quad (9)$$

where the vector $\mathbf{x}_i$ contains the relevant predictors of subject *i* for the probability of giving a non-zero response.

A multivariate zero-inflated Poisson–Gamma model is represented by Eqs. (1), (4), (5), (8) and (9). For discrete time observations, a multivariate zero-inflated Poisson model is represented by Eqs. (1) and (6)–(9). In the fully Bayesian model, priors are required for the parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\alpha}$. The recommended normal priors will be used for $\boldsymbol{\mu}$ and $\boldsymbol{\alpha}$ (Gelman et al., 2004).

The usual prior for the covariance matrix $\boldsymbol{\Sigma}$ is the inverse-Wishart distribution, but the variances of the rates are not identified due to the mixture probabilities ϕ_i. This follows from the expression of the marginal mean,

$$\begin{aligned} E(Y_i^f) = E\left(E\left(Y_i^f \mid \lambda_i^{(f)}\right)\right) &= \phi_i\left(\exp(\mu_1 + \Sigma_{11}/2)\right), \\ &= \exp(\log(\phi_i)\mu_1 + \log(\phi_i)\Sigma_{11}/2), \end{aligned} \quad (10)$$

where Σ_{11} is the first diagonal element of $\boldsymbol{\Sigma}$. Note that the conditional expected rate equals the expected value of a log-normally distributed variable (Aitchison and Ho, 1989). For the multivariate zero-inflated Poisson model, it can be seen that, besides the variance term Σ_{11}, the mixture probability ϕ_i influences the marginal mean and simply serves as a scaling factor. The indeterminacy outlined can be solved by fixing Σ_{11} to an arbitrary constant. In the same way, the variance term Σ_{22} is restricted to identify the scale of the Poisson rate, $\lambda_i(t)$, which represents the expected feedback time. For the Gamma distribution, it can be shown in a similar way that there is an indeterminacy between the variance term Σ_{22} and the mixture probability ϕ_i.

Due to this identification issue, variance terms Σ_{11} and Σ_{22}, representing the between-subject prior variance in the rates, will be restricted to one. The prior for the covariance parameter Σ_{12} to model the correlation between subject's rates will be conditionally specified. Following Klein Entink et al. (2009), this parameter will have a normal prior in the regression of $\log\lambda_i^{(f)}$ on $\log\lambda_i^{(t)}$. Obviously, observed data will be used to assess the between-individual posterior variation.

Note that the model could also be identified by restricting the mean parameter μ_1. However, in that case the mean structure is constructed by the variance components, which can complicate the estimation procedure and the interpretation of the variance components.

When ignoring the zero-inflated component and treating the times as count data, the joint Poisson model resembles the multivariate Poisson-log normal model of Aitchison and Ho (1989). This multivariate Poisson-log normal model can describe the correlation between multivariate discrete count data, where the assumption of independence among the counts can be tested since it is a constrained version of the model. In this field, Böckenholt et al. (2003) proposed a factor-analytic Poisson model, where multiple factors are used to model the dependency structure. Wang (2010) defined a linear model for the Poisson rates using a log link function and latent person variable. The multiple counts are considered to be responses to items, which are indicators of an underlying latent variable.

The multivariate zero-inflated Poisson model allows for overdispersion, since the marginal variance of each count variable, Y_i^f and T_i^f, is greater than the marginal mean relative to an inflated Poisson distribution. The multivariate log-normally distributed Poisson rates induce this overdispersion for the marginal distributions (Aitchison and Ho, 1989). Furthermore, the Poisson rates can be positively and negatively correlated in contrast to other multivariate discrete distributions such as multinomial or negative multinomial (Tunaru, 2002).

3. Latent explanatory variables

Individual differences in the use of feedback and the time spent reading the feedback can be explored through individual explanatory information. In the present computer-based formative assessment the feedback pages can be consulted as an additional source of information after the test has been completed. The results from the assessment are used to collect information about students' abilities and speed of working. Therefore, obvious explanatory variables for student feedback behavior are the skills measured by the test and the working speed. The individual item responses and response times are indicators of the underlying latent variables ability and speed, respectively. The subject-specific latent variables will be used as predictors of the expected feedback use and expected feedback time, and they will constitute the matrix **x** of explanatory variables in Eq. (9).

The response time item response model (RTIRT) will be used to measure the latent variables speed and ability given continuous item response times and discrete item responses (van der Linden, 2007; Fox et al., 2007; Klein Entink et al., 2009). Let Y^a_{ik} and T^a_{ik} denote the dichotomous response and continuous response time of subject i to item k, respectively. At the level of observations, the two-parameter response model is specified as

$$P\left(Y^a_{ik} = 1 \mid \theta_i, a_k, b_k\right) = \Phi\left(a_k\theta_i - b_k\right), \tag{11}$$

and the log-normal response time model as

$$\log T^a_{ik} \mid \zeta_i, c_k, d_k \sim \mathcal{N}\left(d_k - c_k\zeta_i, \omega^2_k\right). \tag{12}$$

At the level of subjects, the subject-specific latent variables are multivariate normally distributed:

$$(\theta_i, \zeta_i)^t \mid \boldsymbol{\mu}_P, \boldsymbol{\Sigma}_P \sim \mathcal{N}_2\left(\boldsymbol{\mu}_P, \boldsymbol{\Sigma}_P\right). \tag{13}$$

In the fully Bayesian framework, priors need to be specified for the item parameters. Following Klein Entink et al. (2009), a multivariate normal prior is defined for the difficulties, $\boldsymbol{b}$, intensities, $\boldsymbol{d}$, the log of discriminations, $\log \boldsymbol{a}$, and the log of time-discriminations, $\log \boldsymbol{c}$. Subsequently, a non-informative normal prior and an inverse-Wishart prior will be used for the mean and covariance matrix, respectively. A different prior model is given by Meyer (2010), who specifies the model with completely independent priors.

The joint distribution of the multivariate outcome data given the latent explanatory variables is given by

$$p\left(\mathbf{y}^f, \mathbf{t}^f \mid \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\zeta}, \boldsymbol{\Sigma}\right) = \prod_i \left[(1-\phi_i) + \phi_i \int_{\mathcal{R}^2_+} p\left(y^f_i, t^f_i \mid \lambda^{(f)}_i, \lambda^{(t)}_i\right) p\left(\log \lambda^{(f)}_i, \log \lambda^{(t)}_i \mid \theta_i, \zeta_i, \boldsymbol{\Sigma}\right) d\boldsymbol{\lambda}\right], \tag{14}$$

where $\mathcal{R}^2_+$ denotes the positive orthant of two-dimensional real space $\mathcal{R}^2$. Note that in this case, interest is focused at the level of subjects, and relationships between feedback (seeking) behavior and latent student characteristics are explored. The latent explanatory student characteristics are related to observable indicators at a lower hierarchical level. This corresponds with the modeling framework of Fox and Glas (2003), where several IRT models were used to measure latent explanatory variables at different levels in a structural multilevel analysis. Skrondal and Rabe-Hesketh (2004, pp. 440–443) describe an application of a hazard model with a latent covariate. Croon and van Veldhoven (2007) and Lüdtke et al. (2008) showed that explanatory variables measured as aggregate scores at the lower level lead to biased regression effects when the associated uncertainty is ignored. Through a latent variable approach, the uncertainty of the higher-level constructs is taken into account.

4. MCMC algorithm

The parameters of both joint models, the multivariate zero-inflated Poisson–Gamma model, Eqs. (1) and (5), and the multivariate zero-inflated Poisson model, Eqs. (1) and (7), will be estimated simultaneously using MCMC. For both models, subject's rates are assumed to be log-normally distributed according to Eq. (8). Furthermore, both models contain a logistic component for the mixture probabilities, Eq. (9), and the RTIRT component for measuring the latent explanatory variables, Eqs. (11) and (12). A sketch is given of the MCMC scheme that represents the full conditional posterior distributions from which samples are to be drawn consecutively. References are given to well-known sampling steps.

The MCMC algorithm consists of several steps to simulate values for the latent explanatory variables and for the parameters of the multivariate zero-inflated model. In iteration m, the first step is defined as the sampling of the latent explanatory values of ability and speed, denoted as $\theta^{(m)}_i$ and $\zeta^{(m)}_i$, respectively. These sampling steps, and the steps for simulating the other RTIRT parameters, are completely described in van der Linden (2007), Klein Entink et al. (2009), and Fox (2010).

Mixture model part

The logistic model parameters $\boldsymbol{\alpha}$ are sampled with a Metropolis–Hastings step. Therefore, consider the random variable Z_i which equals one when feedback is consulted and zero otherwise. For the multivariate Poisson model, Z_i is Bernoulli distributed with a success probability of using feedback

$$\begin{aligned} P\left(Z_i = 1 \mid \boldsymbol{\lambda}_i, \boldsymbol{\alpha}\right) &= P\left(G_i = 1 \mid \boldsymbol{\alpha}\right) P\left(Y^f_i > 0, T^f_i > 0 \mid \lambda^{(f)}_i, \lambda^{(t)}_i, Y^f_i \le K\right) \\ &= \frac{\exp(\mathbf{x}^t_i\boldsymbol{\alpha})}{1+\exp(\mathbf{x}^t_i\boldsymbol{\alpha})}\left(1-\exp\left(-\lambda^{(f)}_i - \lambda^{(t)}_i\right)/C_i(K)\right), \end{aligned} \tag{15}$$

where $C_i(K)$ is the truncation probability defined in Eq. (3). For the multivariate Poisson–Gamma model, the success probability is described by the logistic model with parameter $\boldsymbol{\alpha}$.

Using a multivariate normal proposal distribution for $\boldsymbol{\alpha}$, the candidate $\boldsymbol{\alpha}^{(m)}$ can be evaluated using the acceptance ratio

$$\frac{p\left(\mathbf{z} \mid \mathbf{x}^{(m)}, \boldsymbol{\lambda}^{(f)}, \boldsymbol{\alpha}^{(m)}\right) p\left(\boldsymbol{\alpha}^{(m)} \mid \mu_\alpha, \Sigma_\alpha\right)}{p\left(\mathbf{z} \mid \mathbf{x}^{(m-1)}, \boldsymbol{\lambda}^{(f)}, \boldsymbol{\alpha}^{(m-1)}\right) p\left(\boldsymbol{\alpha}^{(m-1)} \mid \mu_\alpha, \Sigma_\alpha\right)}, \tag{16}$$

where in iteration m, the vector $\mathbf{x}_i^{(m)}$ contains also latent explanatory variables ability and speed values, $\theta_i^{(m)}$ and $\zeta_i^{(m)}$, respectively.

Multivariate zero-inflated (Poisson–Gamma) log-normal part

The rate parameters are also sampled using Metropolis–Hastings. Using the prior defined in Eq. (8), the conditional prior for $\lambda_i^{(f)}$ can be specified as

$$\log \lambda_i^{(f)} \mid \lambda_i^{(t)}, \boldsymbol{\mu}, \Sigma_{12} \sim \mathcal{N}\left(\mu_1 + \Sigma_{12}\left(\log \lambda_i^{(t)} - \mu_2\right), 1 - \Sigma_{12}^2\right), \tag{17}$$

where Σ_{12} is the covariance between the two subject rates, and the variance of each rate is fixed to one. The log posterior distribution of the Poisson rate can be expressed as

$$\begin{aligned} \log p\left(\lambda_i^{(f)} \mid y_i^f, z_i\right) \propto & \left(\log(1-\phi_i) + \phi_i \exp\left(-\lambda_i^{(f)}\right)/C_i(K)\right) I\left(z_i = 0\right) \\ & + \left(\log(\phi_i) - \lambda_i^{(f)} + y_i \log \lambda_i^{(f)} - \log y_i! - \log(C_i(K))\right) I\left(z_i = 1\right) \\ & + \log p\left(\log \lambda_i^{(f)} \mid \lambda_i^{(t)}, \boldsymbol{\mu}, \Sigma_{12}\right). \end{aligned} \tag{18}$$

Proposals are sampled from a normal distribution, and the acceptance ratio can be defined using the posterior density of $\lambda_i^{(f)}$ in Eq. (18). When modeling the times with a Poisson distribution, a Metropolis–Hastings step for the Poisson rate $\lambda_i^{(t)}$ can be defined in a similar way. When using the Gamma distribution, the log posterior distribution of the rate $\lambda_i^{(t)}$ can be expressed as

$$\begin{aligned} \log p\left(\lambda_i^{(t)} \mid t_i^f, z_i\right) \propto & \left(\log(1-\phi_i)\right) I\left(z_i = 0\right) + \left(\log(\phi_i) + \frac{\nu - 2}{2}\log(t_i) - \frac{t_i \nu}{2\lambda_i^{(t)}}\right. \\ & \left. - (\nu/2)\log(2\lambda_i^{(t)}/\nu) - \Gamma(\nu/2)\right) I\left(z_i = 1\right) + \log p\left(\log \lambda_i^{(t)} \mid \lambda_i^{(f)}, \boldsymbol{\mu}, \Sigma_{12}\right), \end{aligned} \tag{19}$$

and a Metropolis–Hastings can be defined using the specified posterior density. The degrees of freedom ν can be sampled from $p(\nu \mid \boldsymbol{\lambda}^{(t)})$, since ν does not enter the likelihood. The posterior can be derived from the posterior $p\left(\boldsymbol{\lambda}^t \mid \nu, \phi, \mathbf{t}, \mathbf{z} = \mathbf{1}\right)$ and a Gamma prior for parameter ν. The posterior density is a non-standard one and Metropolis–Hastings is used to draw samples.

The mean part of the multivariate model can be explicitly modeled such that $\boldsymbol{\mu}_i = \boldsymbol{x}_i^t \boldsymbol{\beta}$, which can include latent covariates. Subsequently, regression parameters $\boldsymbol{\beta}$ in the multivariate lognormal model can be directly sampled given the subject rates and a conjugate multivariate normal prior.

The diagonal elements of $\boldsymbol{\Sigma}$ are restricted to one. Following the approach of McCulloch et al. (2000), the covariance parameter Σ_{12} is a regression parameter in Eq. (17), where the diagonal elements of $\boldsymbol{\Sigma}$ are restricted to one. Let the residual variance parameter be denoted as $\sigma_\rho^2 = 1 - \Sigma_{12}^2$. The residual variance is restricted to be positive, which restricts the parameter space of Σ_{12} to the interval $(-1, 1)$. Assume a normal prior restricted to this interval with mean zero and variance Σ_0. Then, the posterior distribution is given by

$$\Sigma_{12} \mid \boldsymbol{\lambda}, \boldsymbol{\mu}, \sigma_\rho^2 \sim \mathcal{N}\left(\frac{\sum_i \tilde{\lambda}_i^{(t)} \tilde{\lambda}_i^{(f)}/\sigma_\rho^2}{\sum_i \tilde{\lambda}_i^{2(t)}/\sigma_\rho^2 + \Sigma_0^{-1}}, \frac{1}{\sum_i \tilde{\lambda}_i^{2(t)}/\sigma_\rho^2 + \Sigma_0^{-1}}\right) I(-1, 1), \tag{20}$$

where $\tilde{\lambda}_i^{(f)} = \lambda_i^{(f)} - \mu_{i1}$ and $\tilde{\lambda}_i^{(t)} = \lambda_i^{(t)} - \mu_{i2}$, and

$$\sigma_\rho^{-2} \mid \boldsymbol{\lambda}, \boldsymbol{\mu}, \Sigma_{12} \sim Ga\left(\sum_i \left(\lambda_i^{(f)} - \Sigma_{12}(\lambda_i^{(t)} - \mu_2)\right)^2 \Big/ 2 + g_1, \frac{n}{2} + g_2\right) \tag{21}$$

where $Ga()$ denotes the gamma density, and using an inverse gamma prior with scale and shape parameters g_1 and g_2, respectively.

Table 1
Simulation study: parameter estimates and 95% HPD regions across 100 simulated data sets for different within-subject correlations.

True Par.	Model $\mathcal{M}_1$		Model $\mathcal{M}_2$		Model $\mathcal{M}_3$	
	Mean	HPD	Mean	HPD	Mean	HPD
$\rho = 0$	0.00	[−0.10, 0.10]				
$\phi = 0.8$	0.80	[0.76, 0.83]	0.81	[0.77, 0.83]	0.81	[0.77, 0.84]
$\mu_1 = 5$	5.01	[4.91, 5.10]	5.58	[5.58, 5.59]		
$\mu_2 = 5$	5.01	[4.92, 5.10]			5.42	[5.41, 5.42]
$\rho = 0.25$	0.26	[0.16, 0.35]				
$\phi = 0.8$	0.80	[0.76, 0.83]	0.80	[0.76, 0.83]	0.79	[0.76, 0.83]
$\mu_1 = 5$	4.99	[4.90, 5.09]	5.45	[5.45, 5.45]		
$\mu_2 = 5$	5.01	[4.92, 5.10]			5.52	[5.51, 5.52]
$\rho = 0.50$	0.50	[0.42, 0.57]				
$\phi = 0.8$	0.80	[0.76, 0.83]	0.80	[0.76, 0.83]	0.80	[0.76, 0.83]
$\mu_1 = 5$	5.01	[4.91, 5.10]	5.50	[5.49, 5.50]		
$\mu_2 = 5$	5.01	[4.91, 5.10]			5.49	[5.48, 5.49]

5. Simulation study: joint modeling multivariate counts

In this simulation study, the parameter recovery of the developed MCMC algorithm is evaluated for different within-subject correlations and for multivariate count outcomes. The parameters of the zero-inflated multivariate log-normal Poisson model, denoted as model $\mathcal{M}_1$, will be compared with the true values to assess the accuracy of the developed estimation method. Furthermore, the estimates will be compared with the estimates of zero-inflated univariate Poisson models that ignore the within-subject correlation. In the zero-inflated univariate Poisson model, a linear term is defined for the Poisson rate and the probability to be in the Poisson process using the log and the logit link function, respectively.

To emulate the real-data study below, for 500 subjects, 100 data sets were simulated for within-subject correlations of $0, 0.25, 0.50$. For the three different correlations, two log Poisson rates were assumed to be multivariate normally distributed with means of five and variances of one. Next, bivariate count data were generated according to the zero-inflated multivariate log-normal Poisson model, where the probability of feedback use was common across subjects with $\phi = 0.80$ such that 20% of the counts were excess zeros.

The simulations were conducted in R using the developed MCMC scheme. For each data set, 20,000 iterations were made, and a burn-in period of 5000 was sufficient according to Geweke, and the Gelman–Rubin statistic. The intercepts α_0, μ_1, and μ_2 were assigned weakly informative normal priors centered at zero. The covariance parameter Σ_{12} is assumed to be normally distributed in the interval $(-1, 1)$ with mean zero and variance one. The parameters of a zero-inflated univariate Poisson model, with a log-linear model for the Poisson rate, were estimated given the first vector of counts and given the second vector of counts, denoted as model $\mathcal{M}_2$ and $\mathcal{M}_3$, respectively.

The results are given in Table 1. The estimates of model $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$ are given in the second, fourth, and sixth column across the 100 simulations, respectively. Furthermore, the average 95% HPD intervals were computed over the 100 simulations. The first column represents the true parameter values. The estimates of the model $\mathcal{M}_1$ are close to the true value for the different correlation values. Although the prior means were set to zero, the parameter estimates do not show shrinkage toward the prior means. The different correlations between the log Poisson rates were also accurately estimated.

For both univariate models the mean parameter estimates showed a bias. The bias is more or less the same across the different correlation values. This follows from the fact that the expected value of the first and of the second vector of outcomes equals $\exp(\mu_1 + \Sigma_{11}/2) = \exp(5.5)$ and $\exp(\mu_2 + \Sigma_{22}/2) = \exp(5.5)$ given non-excess zeros (see Eq. (10)), respectively, which explains the bias. In this situation of overdispersion the variance in the observed data is also much higher relative to the Poisson distribution. It follows that the univariate Poisson model fails to describe the correlation between the multivariate counts, since it assumes independent Poisson counts. Furthermore, the expected counts are overestimated and the variability in the counts underestimated.

6. Model checking and comparison

Posterior predictive checks will be used to assess the fit of the multivariate zero-inflated models to the data. The posterior predictive data under the model are compared to the observed data using a discrepancy measure. The discrepancy measure takes an extreme value if the model is in conflict with the observed data. For the zero-inflated Poisson model, three different discrepancy measures are often considered to evaluate the fit of the Bernoulli and the Poisson components (e.g., Neelon et al., 2010; Wang et al., 2007).

Hjort et al. (2006) showed that the posterior predictive p-value may not be uniformly distributed under the hypothesized model, which complicates the interpretation of the posterior predictive p-value. They suggested a calibration procedure to correctly interpret the p-values, which requires an additional simulation procedure. Here, the posterior predictive checks are used as a diagnostic tool to identify possible model misfits. This avoids an additional simulation procedure but it is advisable to verify and investigate model misfits suggested by the p-values. That is, a careful investigation of the possible misfit between the observed data and the posited model is needed for situations considered extreme by the p-values.

First, the number of observations equal to zero will be compared to the predicted number of zeros under the model,

$$P\left(\sum_i y_i^{(rep)} = 0 > \sum_i y_i = 0 \mid \mathbf{y}\right),$$

where $y_i^{(rep)}$ are the replicated data under the model. A posterior probability close to one indicates an overestimation, while a posterior probability close to zero indicates an underestimation of the observed amount of zeros.

Second, the index of overdispersion will be used to check whether the model predicts more or less variability in the data relative to the variance of the Poisson model. The posterior predictive check of the overdispersion index is given by

$$P\left(Var(\mathbf{y}^{(rep)})/E(\mathbf{y}^{(rep)}) > Var(\mathbf{y})/E(\mathbf{y}) \mid \mathbf{y}\right).$$

In the case of overdispersion, when the posterior probability is close to one, the model will describe more variability in the data than the Poisson model. The random effect Poisson model component is likely to predict more variance than expected under the Poisson model. Typically, the random effect approach is introduced to accommodate relatively low and high feedback-use observations, and subsequently to account for any overdispersion.

Third, the predictions of the non-zero counts are compared to the observed values. Therefore, the marginal posterior predictive counts are compared with the observed counts, where the posterior predictive check is given by

$$P\left(\sum_i y_i^{(rep)} = c > \sum_i y_i = c \mid \mathbf{y}\right), \quad \text{for } c = 1, 2, \ldots$$

For each category, a posterior probability is computed to check whether the predictions are more extreme than the observed value.

The deviance information criterion (DIC) is used to compare different models. The deviance of the zero-inflated model cannot be directly computed. Therefore, twice the negative log likelihood is used to define the deviance function. Let $\mathbf{\Omega}$ denote the set of relevant model parameters given the observed data $\mathbf{y}$ and $\mathbf{t}$. The deviance is defined as

$$D(\mathbf{\Omega}) = \begin{cases} -2\sum_i \left[\log\left((1-\phi_i) + \phi_i p\left(y_i = 0 \mid \lambda_i^{(f)}, \phi_i\right)\right) + \log(1-\phi_i)\right], & \text{for } y_i = 0, t_i = 0 \\ -2\sum_i \left[\log\left(\phi_i p\left(y_i \mid \lambda_i^{(f)}, \phi_i\right)\right) + \log\left(\phi_i p\left(t_i \mid \lambda_i^{(t)}, \nu, \phi_i\right)\right)\right], & \text{for } y_i > 0, t_i > 0. \end{cases}$$

The DIC is computed as the sum of the posterior mean difference, $\overline{D(\mathbf{\Omega})}$, and the effective number of parameters, p_D. The effective number of parameters is the difference between $\overline{D(\mathbf{\Omega})}$ and the estimated deviance given the posterior mean estimate $\widehat{\mathbf{\Omega}}$.

7. Exploring student feedback behavior

In a computer-based formative assessment (CBFA) at a Dutch university of applied sciences, the opportunity to consult feedback was provided to first-year bachelor students of Law, Health, and Business Administration. After completing the test, the CBFA presented an automatically generated knowledge of results page (summary of correct and incorrect responses) with for each item a link to an additional feedback page. For each item, the additional feedback contains the correct answer and summarized information related to the item content. A student is said to use feedback when he/she visited an additional feedback page. The feedback time was measured as the time between opening and closing the pop-up page showing the additional feedback for that item. For each student and each item, the CBFA registered the response and the response time. Furthermore, the observation whether feedback was consulted and the time taken to process the feedback was also administered.

A CBFA of eleven items on information literacy was used to obtain observational (item) response data on feedback seeking behavior. The eleven items were selected in correspondence with lecturers involved. A total of 610 students completed the test. A brief instruction about the test was given at the start of the assessment, and a written instruction was also handed out. Most of the students completed the test within 45 min, and after this time students were allowed to leave the examination room. The objects are to examine student-specific feedback seeking behavior, including the propensity of using feedback, and the relationship with student-level latent predictors as ability and speed of working.

The total number of feedback pages opened and the total time (in seconds) to process the feedback are the multivariate outcome data. The complexity of the data, with multivariate data and a large number of zeros, requires a multivariate advanced zero-inflated modeling approach. Two multivariate modeling approaches will be considered. In one approach, the total processing times (stored in seconds) are treated as counts using the Poisson distribution, and in the other approach, as positive continuous observations using the Gamma distribution. A Gamma hyperprior was specified for the shape parameter, with shape hyperparameter one and scale hyperparameter five.

The MCMC algorithm was implemented in R (R Development Core Team, 2010). The MCMC algorithm was ran for 100,000 iterations with the first 20,000 iterations as a burn-in. The common convergence diagnostics such as Geweke's and Gelman

Table 2
Feedback behavior study: posterior mean estimates and 95% HPD intervals from empty univariate and multivariate zero-inflated models.

Component	Model $\mathcal{M}_1$		Model $\mathcal{M}_2$		Model $\mathcal{M}_3$	
	Mean	HPD	Mean	HPD	Mean	HPD
Feedback Use (Bernoulli part)						
Intercept, α_0	0.29	(0.13, 0.46)	0.33	(0.17, 0.50)	0.29	(0.13, 0.45)
$1 - \phi$	0.43	(0.39, 0.47)	0.42	(0.38, 0.46)	0.43	(0.38, 0.47)
Feedback Behavior (Poisson–Gamma)						
Intercept, μ_1	1.13	(0.99, 1.26)	1.39	(1.34, 1.45)		
Intercept, μ_2	1.92	(1.81, 2.02)			2.23	(2.16, 2.31)
Correlation, Σ_{12}	0.93	(0.82, 0.99)				
Feedback Behavior (Full Poisson)						
Intercept, μ_1	1.07	(0.97, 1.23)				
Intercept, μ_2	1.90	(1.78, 2.02)			2.69	(2.54, 2.86)
Correlation, Σ_{12}	0.96	(0.85, 0.99)				

and Rubin's did not show any indication of nonconvergence for any of the MCMC chains. The trace plots and convergence diagnostics were evaluated in R.

Estimates of different empty zero-inflated models are given in Table 2. In model $\mathcal{M}_1$, the total number of feedback pages and feedback times were jointly modeled, where the component for the feedback times consisted of a zero-inflated Gamma distribution, according to Eqs. (4) and (5). Additional estimates are reported referred to as full Poisson, where one component models the feedback times using a zero-inflated Poisson distribution, according to Eqs. (6) and (7). Under the label $\mathcal{M}_2$ and $\mathcal{M}_3$, estimates are given of a univariate zero-inflated Poisson (number of feedback pages as outcomes) and a univariate zero-inflated Gamma model (the number of feedback times as outcomes), respectively. Finally, the feedback times were also modeled using a zero-inflated Poisson model, where the estimates are given in the last row of Table 2 under model $\mathcal{M}_3$.

Given the estimates of model $\mathcal{M}_1$, it follows that around 43% of the students did not open any of the feedback pages, and completely ignored this option in the CBFA. The estimated population average equals 57%, which reflects the probability of using feedback. This reflects the high percentage of extra zeros and the requirement of using a zero-inflated Poisson model for the multivariate data. From the estimated log rates follow that on average the expected number of feedback pages visited equals $\exp(1, 13) \approx 3.09$, and the expected total feedback reading time (in seconds) $\exp(1, 92) \approx 6.82$. These estimates are slightly smaller than the average observed values of feedback pages visited and total feedback time, when ignoring the zero observations, which are 4.11 and 9.35, respectively. The within-subject correlation between the log rates is high and around 0.93, which means that the expected total reading time is highly related to the expected number of pages visited. When treating the feedback times as counts, and fitting a multivariate zero-inflated Poisson model, denoted as full Poisson, slightly smaller estimates are obtained.

In Table 2, under model $\mathcal{M}_2$ and $\mathcal{M}_3$, the results are shown of a univariate zero-inflated Poisson model for the total number of feedback pages opened, and a univariate zero-inflated Gamma model for the total feedback reading time. It can be seen that the percentage of extra zeros in both outcomes are comparable with the estimates of 42% and 43%, respectively. The independently estimated Poisson and Gamma mean are slightly higher than the jointly estimated means. It follows that the multivariate normal prior induces more shrinkage toward the general prior mean of zero than the non-informative independent normal prior for the Poisson and Gamma rate in model $\mathcal{M}_2$ and $\mathcal{M}_3$, respectively. When modeling the total times as count data using the zero-inflated Poisson model, estimates comparable to the zero-inflated Gamma model were obtained.

The model fit was investigated of the different zero-inflated models. Given the posterior predictive p-values, all zero-inflated models were capable of predicting adequately the number of extra zeros. The multivariate zero-inflated models slightly underestimated the feedback-use counts of four and five and slightly overestimated the higher counts nine and ten. This is not very surprising, since just two and four students evaluated nine and ten feedback pages, respectively. The posterior predictive check based on the index of overdispersion showed that the Poisson–Gamma model predicts more variability in the observed feedback-use data relative to the variance of the Poisson model (p-value equals 0.97). This shows that the variability of the feedback-use data is much higher than predicted by a Poisson model, which justifies the use of the random effect Poisson component (see, Albert, 1992).

7.1. Exploring student feedback behavior: using background information

Attention will be paid to the multivariate zero-inflated Poisson–Gamma model to analyze the mixed data about the use of feedback, the total number of visited feedback pages, and the total time processing the feedback. However, as part of the joint model, the RTIRT model component (Eqs. (11) and (12)) is used to describe the response data as a function of student's ability and speed of working given the eleven item scores on information literacy. Within the proposed MCMC algorithm for the joint model, posterior draws are made for the latent predictors, which are used to sample the other model parameters.

Table 3
Feedback behavior study: posterior mean estimates and 95% HPD intervals from univariate and multivariate zero-inflated models.

Component	Model $\mathcal{M}_4$		Model $\mathcal{M}_5$		Model $\mathcal{M}_6$	
	Mean	HPD	Mean	HPD	Mean	HPD
Feedback Use (Bernoulli part)						
Intercept, α_0	0.29	(0.17, 0.51)	0.37	(0.20, 0.54)	0.30	(0.14, 0.47)
$1-\phi$	0.42	(0.38, 0.46)	0.41	(0.35, 0.47)	0.43	(0.36, 0.49)
Ability, α_1	0.63	(0.36, 1.10)	0.78	(0.42, 1.14)	0.67	(0.31, 0.98)
Speed, α_2	−0.93	(−1.38, −0.46)	−0.94	(−1.42, −0.48)	−0.88	(−1.36, −0.48)
Feedback Behavior						
Feedback (Poisson part)						
Intercept, β_0	1.15	(0.98, 1.28)	1.39	(1.33, 1.44)		
Ability, β_1	−0.39	(−0.67, −0.13)	−0.36	(−0.48, −0.25)		
Speed, β_2	−0.26	(−0.59, 0.13)	−0.15	(−0.30, −0.02)		
Feedback-Time (Gamma part)						
Intercept, β_0	1.93	(1.80, 2.03)			2.70	(2.64, 2.75)
Ability, β_1	−0.25	(−0.48, −0.03)			−0.86	(−0.95, −0.78)
Speed, β_2	−0.36	(−0.65, −0.05)			−0.60	(−0.71, −0.53)
Correlation, Σ_{12}	0.93	(0.81, 0.99)				

Subsequently, in the full conditional joint model, the latent explanatory variables working speed and ability are used as predictors.

The predictors are used to explore differences in the use of feedback through a logistic model. That is, the mixture probabilities are conditionally modeled as

$$\phi_i = P\left(G_i = 1 \mid \boldsymbol{\alpha}, \theta_i, \zeta_i\right) = \frac{\exp(z_i)}{1+\exp(z_i)},$$
$$z_i = \alpha_0 + \alpha_1\theta_i + \alpha_2\zeta_i,$$

such that student-specific ability and speed are related to the probability of using feedback. Subsequently, student response rates of feedback (i.e. expected number of feedback pages visited and expected total feedback reading time) are multivariate log-normally distributed,

$$\log\lambda_i^{(f)} = \beta_{00} + \beta_{01}\theta_i + \beta_{02}\zeta_i + \eta_{0i},$$
$$\log\lambda_i^{(t)} = \beta_{10} + \beta_{11}\theta_i + \beta_{12}\zeta_i + \eta_{1i}$$

where $\boldsymbol{\eta}_i$ is multivariate normally distributed with variances restricted to one and correlation parameter Σ_{12}. Normal independent priors for the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are specified with zero means and large variances.

In Table 3, the estimated parameters are given of the full joint model, denoted as model $\mathcal{M}_4$. A univariate zero-inflated Poisson and a univariate zero-inflated Gamma model with explanatory variables are considered to model the total number of visited feedback pages and the total feedback time, denoted as model $\mathcal{M}_5$ and $\mathcal{M}_6$, respectively.

From model $\mathcal{M}_4$ follows that the average percentage of students not visiting any feedback page is around 42%, which was also found with the empty joint models (see Table 2). The student-specific predictors ability and speed of working have a significant relationship with the probability of using feedback. It follows that students who performed well on the test have a significantly higher probability of using feedback (i.e. visiting at least one page) than those who performed poorly. For an average working speed, an increase in ability of 0.50 leads to an increase of around 8% in the probability of using feedback. High ability students were specifically interested in feedback to incorrectly answered questions, whereas low ability students were probably less motivated to check out the feedback. The working speed is negatively related with the probability of using feedback. This means that students who work fast have a lower probability of using feedback than students who work more slowly. For an average ability, an increase of 0.50 in working speed leads to an average decrease of 11% in the probability of using feedback.

In Fig. 1, the estimated student-specific probabilities of using feedback are plotted against the student ability and speed of working. It follows that variations in ability and speed of working lead to differences in the probability of using feedback, where high ability students who worked slowly on the test have a high chance of using feedback. Students who worked very fast and those who performed poorly on the test have a low chance of using the feedback. Here, it seems that test motivation is an important factor, since students who took the test seriously, who performed well and took their time, also visited feedback pages.

The estimates of the multivariate Poisson–Gamma part show negative effects for both predictors. There is a negative relation between ability and speed of working and the expected total number of visited feedback pages. Although high ability students are more likely to visit a feedback page than low ability students, they are expected to visit less feedback pages in total. For a student with an average working speed, high ability students ($\theta = 1$) visit on average two feedback pages (i.e., approximately $\exp(1.15-0.39)$), low ability students ($\theta = -1$) on average four to five feedback pages and students of average ability ($\theta = 0$) on average three feedback pages. This corresponds to the finding that high ability students consult

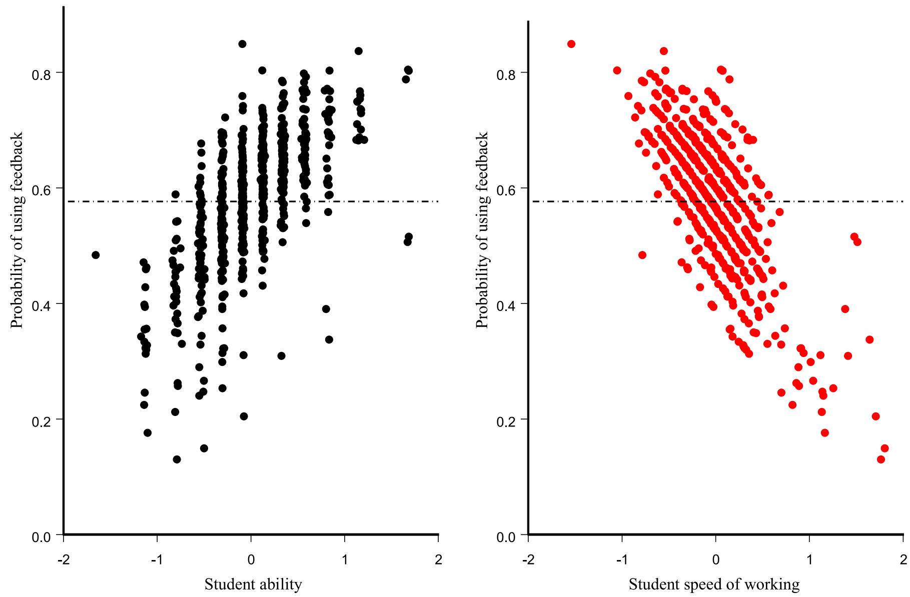

Fig. 1. Posterior probability of using feedback as a function of student ability and speed of working.

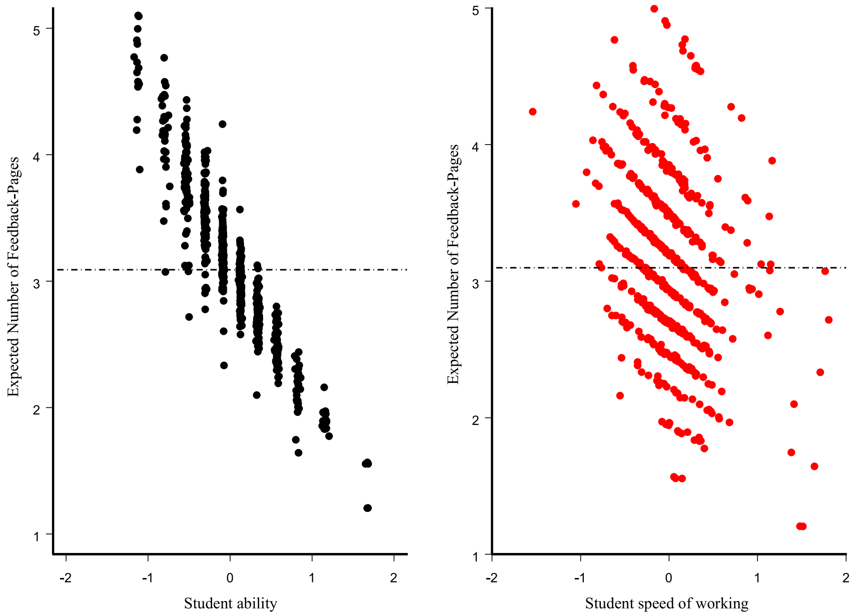

Fig. 2. Posterior expected total number of visited feedback pages as a function of student ability and speed of working.

feedback of the few items that were made incorrectly. The working speed has a negative effect on the total number of pages visited but this effect is not significant.

In Fig. 2, the posterior expected total number of feedback pages is plotted against ability and speed of working using the estimates from model $\mathcal{M}_4$. It can be seen that high ability students visited one to two feedback pages, whereas low ability students visited four to five pages. On average slightly more than three pages were visited. It follows that the total number of visited pages varies a lot for students with the same working speed.

Finally, the total time processing the feedback is negatively related to ability and working speed. The estimated effects are such that the average feedback processing time is around 6.8, where low ability students spend more time reading feedback and high ability students less time. The same pattern is found for fast and slow working students. In Fig. 3, the posterior expected total time is given for each student. The negative trend between student ability and total time shows that high ability students read a few specific feedback pages, whereas others consult more pages and use more reading time. The negative trend between the total time and speed of working suggests that fast working students also work more quickly

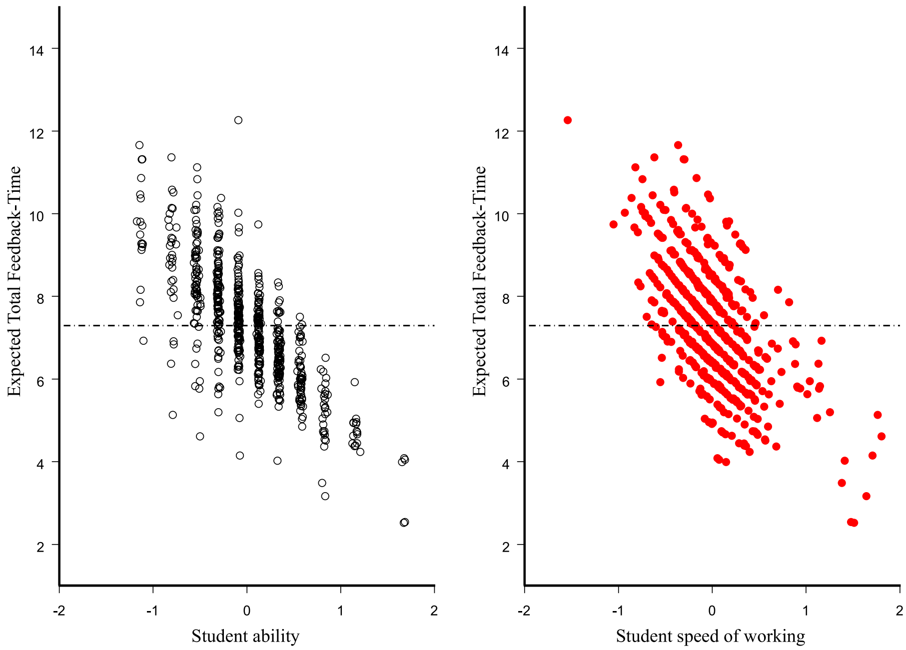

Fig. 3. Posterior expected total time reading the feedback as a function of student ability and speed of working.

Table 4
Feedback behavior study: multivariate zero-inflated model comparison.

Model	Deviance	p_D	DIC
$\mathcal{M}_1$			
Full Poisson	4396.75	268.78	4665.53
Poisson–Gamma	4113.14	346.06	4459.20
$\mathcal{M}_4$			
Full Poisson	4313.24	284.03	4597.27
Poisson–Gamma	4048.39	362.09	4410.38

through the feedback pages. However, the fast working students were also known to be less motivated to use feedback (see Fig. 1), which may also lead to less processing time for those who consulted feedback. They could have opened a feedback page and closed it quickly without paying much attention to the content.

The log rates are highly positively correlated, around 0.93. From the results of model $\mathcal{M}_5$ and $\mathcal{M}_6$ follows that the effects of the predictors are quite similar. However, as in the empty joint model, the intercepts of the independently estimated rates are slightly higher compared to the jointly estimated intercepts. The correlation between the log rates leads to slightly smaller expected values. That is, the total number of consulted feedback pages is expected to be smaller given the total feedback time in comparison to the unconditional expected value. In the same way, the total feedback time is expected to be smaller given the total number of consulted pages than the unconditional expected value. For the Gamma component, the estimated degrees of freedom ν equals 38.78 and 40.16 for the empty and the conditional model, respectively.

Finally, results of a small model comparison study are given in Table 4. The empty multivariate zero-inflated Poisson–Gamma model fits the data better than the full Poisson model, where the feedback times are considered proxies and treated as counts. It follows that the random effect Gamma component has more free parameters, which leads to a substantially better fit. The fit of the model is further improved by incorporating the explanatory variables, according to the DICs of model $\mathcal{M}_4$. However, from the DICs follow that the empty Poisson–Gamma model fits the data better than the full Poisson model with explanatory variables. This shows that the simplifying case of modeling multivariate counts instead of modeling outcomes in different scales leads to less optimal results with respect to model fit.

8. Discussion

A multivariate zero-inflated Poisson–Gamma model has been proposed to explore student feedback behavior in a computer-based assessment from which counts and continuous data were observed. The proposed joint model consists of a binary component and a multivariate component. The binary component models the use of feedback. The multivariate component models the total number of feedback pages consulted using a truncated Poisson distribution and the total feedback processing time using the Gamma distribution. At a higher-level, the student-specific log means or rates of both components are assumed to be normally distributed. A multivariate zero-inflated Poisson model is given for the situation where both observations are discrete measurements. In a simulation study it is shown that the developed MCMC algorithm can recover accurately all model parameters and that modeling the count data separately with a zero-inflated univariate Poisson model leads to parameter bias in situations of overdispersion. It is shown that the multivariate model can provide accurate random effect Poisson and Gamma rate predictions, which also leads to accurate predictions of counts and other quantities of interest.

Further, it is shown that latent explanatory variables (ability and speed) can be integrated using an RTIRT measurement model. Therefore, posterior draws of the latent explanatory variables are used to sample the other model parameters.

The proposed joint model has several important features. First, it addresses the extra zeros through a mixture component relative to the ordinary Poisson and Gamma model. Second, it models the use of feedback as well as the feedback intensity as the total number of pages consulted and the total reading time. Third, the joint modeling approach accounts for dependence between model components, which can lead to less biased inferences. Fourth, it can handle within-subject correlation between observed mixed outcomes. Fifth, explanatory information can be incorporated to model heterogeneity in the probability of feedback use and, given feedback use, feedback behavior.

The posterior predictive checks showed that the use of log-normally distributed rates improved the model's capability to describe the variability in the data. However, extreme discrete outliers were difficult to model correctly. Further research will focus on a normal mixture prior for the rates to capture extreme cases, where subjects with extreme observations can also be classified as such. However, such a nested structure of mixtures is complex, where the normal mixture prior is nested within the mixture component to handle zero-inflated count data, and requires proper identification rules (Vermunt, 2003; Vermunt and Magidson, 2005).

The purpose of the Dutch feedback behavior study was to explore feedback behavior in a CBFA. Furthermore, the object was to explain individual heterogeneity in their willingness to seek feedback using student-characteristic variables. It was shown that heterogeneity in feedback use was partly explained by student achievement and speed of working. Student achievement also explained a significant amount of variation in the expected number of pages consulted and the expected total reading time. The student working speed did show a less clear relationship with the expected outcomes of feedback behavior. Various studies show different effects of feedback in computer-based assessment on student learning, where the effects are influenced by individual differences in feedback behavior and ways of providing feedback. In specific, student differences in feedback seeking behavior complicate a straightforward efficient implementation of feedback in a computer-based assessment. For example, Timmers and Veldkamp (2011) showed that the attention paid to feedback differs over students, where student characteristics such as study motivation and a positive attitude influence the time spent reading feedback.

Different reasons can be given for the relatively low percentage of students using feedback. First, students were given feedback regarding their performances on a low-stake test, which did only motivate the high ability students to seek feedback. Second, the feedback was provided after the test. The feedback use might increase when providing the feedback immediately after responding to an item. In this context it is also possible to consider system-initiated feedback such that feedback can be provided automatically after an incorrect answer is given.

This study shows that there is high degree of between-individual heterogeneity in feedback use. Although the relation between feedback and learning outcomes receives much attention, individual variability in feedback use is generally neglected. To study accurately the effects of feedback interventions on learning outcomes this variation in feedback use should be taken into account.

Acknowledgment

The author thanks Caroline Timmers for providing the data from the computer-based assessment for learning on information literacy.

References

Aitchison, J., Ho, C.H., 1989. The multivariate Poisson-log normal distribution. Biometrika 76, 643–653.

Albert, J., 1992. A Bayesian analysis of a Poisson random effects model for home run hitters. The American Statistician 4, 246–253.

Böckenholt, U., Kamakura, W.A., Wedel, M., 2003. The structure of self-reported emotional experiences: a mixed-effects Poisson factor model. British Journal of Mathematical & Statistical Psychology 56, 215–229.

Congdon, P., 2005. Bayesian Models for Categorical Data. Wiley, Chichester.

Croon, M.A., van Veldhoven, M.J.P.M., 2007. Predicting group-level variables from variables measured at the individual level: a latent variable multilevel model. Psychological Methods 12, 45–57.

Fox, J.P., 2010. Bayesian Item Response Modeling: Theory and Applications. Springer, New York.

Fox, J.P., Glas, C.A.W., 2003. Bayesian modeling of measurement error in predictor variables using item response theory. Psychometrika 68, 169–191.

Fox, J.P., Klein Entink, R.E., van der Linden, W.J., 2007. Modeling of responses and response times with the package cirt. Journal of Statistical Software 20 (7).

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis, second ed. Chapman & Hall, New York.

Hattie, J., Timperly, H., 2007. The power of feedback. Review of Educational Research 77, 81–112.

Hjort, N.L., Dahl, F.A., Steinbakk, G.H., 2006. Post-processing posterior predictive *p*-values. Journal of the American Statistical Association 101, 1157–1174.

Hu, M.C., Pavlicova, M., Nunes, E.V., 2011. Zero-inflated and hurdle models of count data with extra zeros: examples from an HIV-risk reduction intervention trial. American Journal of Drug and Alcohol Abuse 37, 367–375.

Klein Entink, R.H., Fox, J.P., van der Linden, W.J., 2009. A multivariate multilevel approach to the modeling of accuracy and speed of test takers. Psychometrika 74, 21–48.
Lambert, D., 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics 34, 1–14.
Loeys, T., Moerkerke, B., De Smet, O., Buysse, A., 2012. The analysis of zero-inflated count data: beyond zero-inflated Poisson regression. British Journal of Mathematical and Statistical Psychology 65, 163–180.
Lüdtke, O., Marsh, H.W., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., 2008. The multilevel latent covariate model: a new, more reliable approach to group-level effects in contextual studies. Psychological Methods 13, 203–229.
McCulloch, C., 2008. Joint modelling of mixed outcome types using latent variables. Statistical Methods in Medical Research 17, 53–73.
McCulloch, R.E., Polson, N.G., Rossi, P.E., 2000. A Bayesian analysis of the multinomial probit model with fully identified parameters. Journal of Econometrics 99, 173–193.
Meyer, J.P., 2010. A mixture Rasch model with item response time components. Applied Psychological Measurement 34, 521–538.
Mihaylova, B., Briggs, A., O'Hagan, A., Thompson, S.G., 2011. Review of statistical methods for analysing healthcare resources and costs. Health Economics 20, 897–916.
Min, Y., Agresti, A., 2005. Random effect models for repeated measures of zeroinflated count data. Statistical Modeling 5, 1–19.
Neelon, B.H., O'Malley, A.J., Normand, S.H., 2010. A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use. Statistical Modelling 10, 421–439.
R Development Core Team 2010. R: A Language and Environment for Statistical Computing. In: R Foundation for Statistical Computing, Vienna.
Rabe-Hesketh, S., Skrondal, A., 2007. Multilevel and latent variable modeling with composite links and exploded likelihoods. Psychometrika 72, 123–140.
Rizopoulos, D., Verbeke, G., Molenberghs, G., 2010. Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival outcomes. Biometrics 66, 20–29.
Rose, E., Martin, S.W., Wannemuehler, K.A., Plikaytis, B.D., 2006. On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. Journal of Biopharmaceutical Statistics 16, 463–481.
Skrondal, A., Rabe-Hesketh, S., 2004. Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models. Chapman & Hall, Boca Raton, Florida.
Timmers, C.F., Veldkamp, B.P., 2011. Attention paid to feedback provided by a computer-based assessment for learning on information literacy. Computers & Education 56, 923–930.
Tunaru, R., 2002. Hierarchical Bayesian models for multiple data. Austrian Journal of Statistics 31, 221–229.
van der Kleij, F.M., Eggen, T., Timmers, C.F., Veldkamp, B.P., 2012. Effects of feedback in a computer-based assessment for learning. Computers & Education 58, 263–272.
van der Linden, W.J., 2007. A hierarchical framework for modeling speed and accuracy on test items. Psychometrika 72, 287–308.
van der Linden, W.J., Glas, C.A.W. (Eds.), 2010. Elements of Adaptive Testing. Springer, New York.
Vermunt, J., 2003. Multilevel latent class models. Sociological Methodology 33, 213–239.
Vermunt, J., Magidson, J., 2005. Hierarchical mixture models for nested data structures. In: Weihs, C., Gaul, W. (Eds.), Classification: The Ubiquitous Challenge. Springer, Heidelberg, pp. 176–183.
Wang, L., 2010. IRT-ZIP modeling for multivariate zero-inflated count data. Journal of Educational and Behavioral Statistics 35, 671–692.
Wang, H.M., Kalwani, M., Akçura, T., 2007. A Bayesian multivariate Poisson regression model of cross-category store brand purchasing behavior. Journal of Retailing and Consumer Services 14, 369–382.
Yang, Y., Kang, J., Mao, K., Zhang, J., 2007. Regression models for mixed Poisson and continuous longitudinal data. Statistics in Medicine 26, 3782–3800.