METRON (2021) 79:57–91
https://doi.org/10.1007/s40300-021-00200-1

The Impact of COVID-19 on Students' Marks: A Bayesian Hierarchical Modeling Approach

Jabed Tomal[1] · Saeed Rahmati[1] · Shirin Boroushaki[1] · Lingling Jin[2,4] · Ehsan Ahmed[3]

Received: 28 August 2020 / Accepted: 4 February 2021 / Published online: 17 February 2021

Abstract
Due to COVID-19, universities across Canada were forced to undergo a transition from classroom-based face-to-face learning and invigilated assessments to online-based learning and non-invigilated assessments. This study attempts to empirically measure the impact of COVID-19 on students' marks from eleven science, technology, engineering, and mathematics (STEM) courses using a Bayesian linear mixed effects model fitted to longitudinal data. The Bayesian linear mixed effects model is designed for this application which allows student-specific error variances to vary. The novel Bayesian missing value imputation method is flexible which seamlessly generates missing values given complete data. We observed an increase in overall average marks for the courses requiring lower-level cognitive skills according to Bloom's Taxonomy and a decrease in marks for the courses requiring higher-level cognitive skills, where larger changes in marks were observed for the underachieving students. About half of the disengaged students who did not participate in any course assessments after the transition to online delivery were in special support.

✉ Saeed Rahmati
srahmati@tru.ca

Jabed Tomal
jtomal@tru.ca

Shirin Boroushaki
sboroushaki@tru.ca

Lingling Jin
ljin@tru.ca; lingling.jin@cs.usask.ca

Ehsan Ahmed
eahmed@tru.ca

1 Department of Mathematics and Statistics, Thompson Rivers University, Kamloops, BC V2C 0C8, Canada

2 Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan S7N 5C9, Canada

3 Department of Architectural and Engineering Technology, Thompson Rivers University, Kamloops, BC V2C 0C8, Canada

4 Department of Computing Science, Thompson Rivers University, Kamloops, BC V2C 0C8, Canada

Keywords COVID-19 · Science education · Hierarchical Linear Modeling · Missing value estimation

1 Introduction

The spread of the novel coronavirus disease 2019 (COVID-19) started in December 2019 in Wuhan, China [32]. Due to the rising health concerns, many universities around the world transitioned from face-to-face to online course delivery and assessments [24]. In the province of British Columbia (BC), Canada, all on-campus classes and activities in the post-secondary schools were cancelled starting March 15, 2020. At Thompson Rivers University (TRU), the face-to-face classes shifted to alternative modes of delivery mostly conducted through the online learning management systems. In order to adapt to the changes, most, if not all, of remaining course assessments were switched to open-book non-invigilated exams with adjusted weights for the evaluation components. In this study, we proposed a Bayesian statistical model to evaluate the effects of the sudden change from classroom-based to online delivery and assessments in response to COVID-19 pandemic on students' academic performance. It is important to understand comprehensively how students' performance has changed after March 15, 2020 due to the COVID-19 effect and to investigate factors that potentially contributed to the changes.

Due to the unique situation of COVID-19 and its social and educational ramifications, there have been a few published articles discussing how COVID-19 and the shutdown of universities have impacted students' performance in post-secondary education. For instance, Sintema [26] investigated the possible impacts that the closure of secondary schools due to COVID-19 in Zambia would have on the general performance of students in specific subject areas. Having interviewed STEM[1] educators at a public secondary school in Zambia, the study concluded that learners' performance would be negatively affected in STEM subjects in the upcoming national examination if the COVID-19 epidemic is not being lessened in the shortest possible time. This is because of not only the lack of contact and meaningful interactions between learners and teachers, but also insufficient e-learning tools to facilitate such interactions. Basilaia and Kvavadze [3] studied the capacities of the country Georgia and its population to continue the education process at the schools in the online form of distance learning. The study was conducted in a private school during the COVID-19 pandemic which reviewed different available platforms that were used with the support of the government for online education and live communications. Their findings showed that the transition from traditional to online education system was successful in terms of adaptability and gained skills by students, teachers and administrative staff. Nevertheless, none of these studies compared students' performances empirically.

During the Severe Acute Respiratory Syndrome (SARS) epidemic in Hong Kong, when all schools and universities were ordered closed and governments invoked quarantine laws to isolate those who might be the carriers, Wong [29] conducted a study which briefly described the impact of e-commerce on the local community with an emphasis on the use of e-learning technology as a contingency measure in tertiary institutions. Their study showed that, given limited time available for the course design and delivery, the examination result of the e-learning class was slightly better than the traditional class. However, their study lacks the

[1] *STEM* is a curriculum based on the idea of educating students in four specific disciplines—Science, Technology, Engineering and Mathematics—in an interdisciplinary and applied approach.

rigorous empirical comparisons of students' marks using a sound statistical model. They suggested further rigorous study.

There are some studies evaluating how students perform in online versus face-to-face delivery in post-secondary mathematics education. Jones and Vena [13] focused on equity in learning as reflected in the final grades of online and on-site students from the same post-secondary mathematics course taught repeatedly over 10 semesters. On-site students attended regular class sessions, while online students only attended an orientation session and a final exam. For both groups, the evaluations were invigilated. Their findings revealed significant statistical differences in online and on-site students' final grades, in favor of on-site student achievement. Rey [23] examined the associations between taking basic skills mathematics courses online versus face-to-face and student success and persistence. The study stressed on the difficulties associated with effective communication of mathematics topics and ideas via the Internet and found no noticeable associations with learning outcomes or persistence. Their study also pointed out that the quality of education gained from online basic skills mathematics courses is relatively equivalent to face-to-face courses. Weems [27] compared two sections of beginning algebra course where one was taught online and the other was taught on site. The study reported no significant difference for exam average between the two formats, but highlighted a decrease in performance by the online students across the exams. Ashby et al. [2] compared student success in a Developmental Math course offered in three different learning environments: online, blended, and face-to-face. Using a one way analysis of variance (ANOVA), the authors showed that there were significant differences between learning environments with the students in the blended courses having the least success. With regards to the efficiency of learning outcomes in the online pedagogy, Arias et al. [1] studied the effectiveness of online delivery relative to face-to-face delivery by randomly assigning students into two categories: online and face-to-face. These two sections were taught by the same instructor and the course objectives and exams were the same for both sections as well. The authors concluded that "both course objectives and the mechanism used to assess the relative effectiveness of the two modes of education may play an important role in determining the relative effectiveness of alternative delivery methods." One other area of interest in this respect is the notion of exams and their competence in the digital space. In a study by Williams and Wong [28], the authors examined the efficacy of closed-book invigilated final exams versus open-book and open-web (OBOW) final exams in a completely online university. They analyzed students' experience who had completed both formats of the exam by surveying them on the merits of each exam format. Their findings showed that 100% of students found OBOW preferable to traditional closed-book exams. On the issue of academic integrity, their results indicated that in both exam formats "there has been equal opportunity of plagiarism in the view of students." On the other hand, there are some studies indicating that online exams and access to technology provide more opportunities for dishonest behaviors [14]. Even though there are some similarities and differences with the aforementioned studies, the research environment of our study is qualitatively different. In our study, a Bayesian statistical model is proposed to compare students' performances reflected by their empirical marks received in a sudden unexpected scenario where both classes and evaluations were virtually forced to be conducted online.

On the other hand, the closure of universities and transitioning to online teaching may have prompted mental, psychological and educational challenges for students and faculty members. A recent study by Statistics Canada based on crowdsourcing data completed by over 100,000 post-secondary students from April 19 to May 1, 2020 provides insight into how students' academic life was impacted by the COVID-19 pandemic [11]. Admittedly, the adoption of alternate modes of delivery and distance learning under these circumstances are not ideal. Students accustomed to on-campus learning have expressed concerns over the loss of social and interactive side of education. In addition, the rapid transition to online instruction poses challenges to students who were not equipped to adjust to this mode of learning either because they learn better in-person, lack appropriate tools, suffer financial hardship, or do not have a home environment suitable for learning online [11]. The transition of classes to online from in-person lectures due to COVID-19 had an even more severe impact on students with mental health disabilities who needed extra care and frequent face-to-face interactions with the instructors in order to keep maintaining motivation, interest and persistence to succeed in the course [9]. In a recent study, Sahu [24] highlighted the impacts of COVID-19 on education and mental health of students and academic staff and indicated the challenges posed by the closure of universities and other restrictive measures such as the shift in the delivery mode of the course, the change in the format of assessments, the travel restrictions on students, etc. In another study, Cao et al. [5] investigated students' anxiety level during COVID-19 in a Chinese medical college located in Hubei Province. Their findings showed that about a quarter of students in their sample had experienced mild, moderate or severe levels of anxiety during COVID-19 outbreak. In this paper, we have studied some aspects of stress among students with disabilities and special needs that contributed to their disengagement in any of the remaining evaluation components after the emergence of COVID-19.

We developed a Bayesian hierarchical linear mixed effects model to measure the effects of COVID-19 on students' marks. The literature of random effects model goes back to Laird and Ware [16]. The classical maximum likelihood estimation and inference of the linear mixed effects model using expectation maximization (EM) algorithm were shown by Laird et al. [15] and Lindstrom et al. [19]. These articles approached the problem of model building and inference from a frequentist viewpoint. On the other hand, the general design of Bayesian method for the linear mixed effects model was described by Zhao et al. [31]. The missing value imputation for linear mixed effects model, using a frequentist viewpoint, was proposed by Schafer et al. [25]. In this paper, we redesigned the linear mixed effects model for the change of students' raw marks before and after COVID-19 effect in Winter 2020 semester at TRU. We described the complete Bayesian methodology for the proposed model using conjugate and semi-conjugate prior distributions. We then derived the full conditional posterior distributions of the parameters and proposed the Gibbs sampling [17] to generate Markov Chain Monte Carlo (MCMC) samples [12] from the posterior distributions. In order to impute missing values of marks, we assumed the mechanism of missingness completely at random [4]. Our novel contributions are the design and implementation of the fully Bayesian missing value imputation method in a linear mixed effects modeling setup. While most of the classical statistical missing value imputation methods are performed only once given complete data before the model building [7], our method of Bayesian missing value imputation is flexible which seamlessly generates missing values before every generation of the Markov chain from the posterior distributions of the missing values given observed data. In our model, we allow student-specific error variances to vary which is a considerable extension of the methodology in a Bayesian linear mixed effects hierarchical modeling setup. We wrote our codes for the proposed fully Bayesian hierarchical model using `R` [22]. The `R` codes and data are available on GitHub (https://github.com/jhtomal/covid19_impact.git). We are in the process of wrapping up the codes in an `R` package to publish in the Comprehensive R Archive Network (CRAN) so that the broad scientific community can apply our method in their applications.

We hypothesize that the COVID-19 has negative effects on students' performances, reflected on their marks, and on their stress level especially on students needing special supports. As such, this article aims to explore the following questions:

(i) How are the raw marks of all students in a course compared before and after the university switched to online delivery due to COVID-19?
(ii) How are the raw marks of individual students within a course compared before and after the transition to online delivery?
(iii) Are the stronger or weaker students getting higher or lower average raw marks due to online delivery relative to in-person delivery?
(iv) Are there disengaged students who did not participate in any evaluation components in a course after the university shifted to online delivery? Are these students on special support?
(v) Are the trends of marks consistent across all courses or departments in this study? If there are different trends, what are the potential factors that make the difference?

In order to cope with the the unexpected transition to online delivery and open-book non-invigilated evaluations, the instructors from different departments in the Faculty of Science at TRU came up with new sets of weighted schemes which were strikingly different from what they had at the beginning of the semester. We thus analyze students' raw marks, as opposed to weighted aggregated marks, because the raw numbers reflect the unbiased scenario. Subsequently in Sect. 5, we discussed the results according to the level of cognitive skills and hands-on experiences required in the courses we investigated. In our analysis, we classified the cognitive skills with reference to the Bloom's Taxonomy of Knowledge.[2]

2 Design of the study and description of data

The analysis of empirical data starts with some summary statistics followed by a fully Bayesian linear mixed effects model. The posterior statistical inferences of the model parameters proceed following posterior model inspections.

The study population consists of all of the students in the Faculty of Science at Thompson Rivers University. We collect longitudinal data by sampling 11 courses (with total number of students equals 326; see Table 1 for details) across three departments from the Faculty of Science, including seven courses in Mathematics (MATH) and Statistics (STAT), two courses in Computing Science (COMP) and two courses in Architectural and Engineering Technology (ARET). Each course has been taught, marked,[3] and graded by the same instructor from the start to the end of the Winter 2020 semester (before and after COVID-19 effects) which ensures instructor's effects being adjusted while leaving only students' effects to compare. Moreover, the Cronbach's alpha [8] test of reliability analysis has been performed to each course to make sure that the course data are reliable. After the selection of the courses, student-specific raw marks are recorded over time using a series of evaluations such as *multiple* assignments, quizzes, tests, exams, and projects. Specific to each course, we have analyzed marks from all sorts of evaluations and ensured covering all possible grading practices. The raw marks specific to each evaluation component are then converted into percents (from 0 to 100) where larger numbers indicate better marks. We also record whether the marks are observed or *missing*. Given that all the classes went online starting March 15, 2020, the index variable is simply defined as a vector of *boolean* values indicating whether the effects of COVID-19 have occurred (before March 15 vs after March 15). As there is no randomness in

2 *Bloom's Taxonomy of Knowledge* is a set of three hierarchical models used to classify educational learning objectives into levels of complexity and specificity: https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/.

3 There is no teaching assistant for marking.

Table 1 Number of students, percentage of disengaged students, and percentage of missing marks

Courses	Number of students	Disengaged students (%)	Missing marks (%)
MATH1070	36	5.56	3.78
MATH1240	29	0	3.73
MATH1250	35	0	5.14
MATH1640	29	3.45	3.86
MATH2200	19	0	6.22
MATH2240	38	0	4.67
STAT2000	29	3.45	6.34
ARET1400	34	2.94	7.67
ARET2600	21	0	13.09
COMP2680	28	3.57	14.07
COMP4980	28	0	5.65

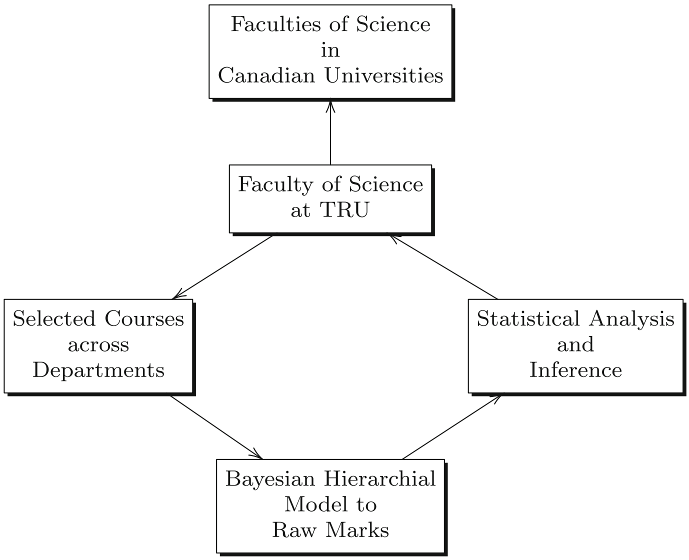


Fig. 1 Overall design of the study

the index variable, we consider it deterministic. A Bayesian hierarchical model with missing value imputation technique is then applied to determine whether students' raw marks were increased or decreased before and after March 15. After model fitting, statistical analysis and inference are performed to student specific raw marks in each course. Our findings generalize to the Faculty of Science at TRU, and may eventually to all faculties of science across universities in Canada. Figure 1 shows the overall design of the study.

3 Methods

3.1 The linear mixed effects model

Let Y_{it} be the raw marks in percent for the ith student ($i = 1, \ldots, k$) evaluated at time t ($t = 1, \ldots, n$) in a particular course. The subscript t represents a series of evaluations conducted over time from *multiple* assignments, quizzes, tests, exams, and projects. Let x_{it} be another variable measured at time t for the ith student which can explain the variation in Y_{it}. We consider the following linear mixed effects model

$$Y_{it} = \beta_{0i} + \beta_{1i} x_{it} + \epsilon_{it}, \tag{1}$$

where

$$x_{it} = \begin{cases} 1 & \text{evaluation taken after March 15, 2020} \\ 0 & \text{otherwise,} \end{cases}$$

and $\epsilon_{it} \sim \text{Normal}(0, \sigma_i^2)$ with student-specific error variance σ_i^2. The sampling distribution of Y_{it} is

$$Y_{it} | \beta_{0i}, \beta_{1i}, \sigma_i^2 \sim \text{Normal}\left(\beta_{0i} + \beta_{1i} x_{it}, \sigma_i^2\right). \tag{2}$$

In this model, β_{0i} and $\beta_{0i} + \beta_{1i}$ are the average marks of the ith student before and after March 15, 2020, respectively. Here, β_{1i} is the difference of average marks before and after March 15, 2020 for the ith student. Let $\boldsymbol{\beta}_i = (\beta_{0i}, \beta_{1i})^T$ represents the vector of regression coefficients for the ith student. We assume that the ith student is a randomly and independently selected student from the pool of students in a specific course in the faculty of science at Thompson Rivers University (TRU). This leads us to consider the sampling distribution for $\boldsymbol{\beta}_i$ as

$$\boldsymbol{\beta}_i | \Sigma \sim \text{Multivariate Normal}\left(\boldsymbol{\theta}, \Sigma\right), \tag{3}$$

where $\boldsymbol{\theta} = (\theta_0, \theta_1)^T$ and Σ is the variance-covariance matrix. This part of the model explains that θ_0 and $\theta_0 + \theta_1$ are the average marks of all students in a course before and after March 15, 2020, respectively.

The sampling distribution for the student-specific error variance is

$$\frac{1}{\sigma_i^2} \sim \text{Gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right), \tag{4}$$

defined in terms of *shape* and *rate* parameters, where the course specific error variance is σ_0^2 with strength ν_0 (> 0). Here, the large values of ν_0 will force $1/\sigma_i^2$ to be tightly clustered around σ_0^2 and vice versa. On the other hand, large values of σ_0^2 represent large between-student variability and small within-student variability σ_i^2. The overall picture of the hierarchical model is shown below in Fig. 2.

The likelihood function for the parameters of the linear mixed effects model is:

$$\begin{aligned} &L\left(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k, \sigma_1^2, \ldots, \sigma_k^2, \boldsymbol{\theta}, \Sigma, \nu_0, \sigma_0^2 | \mathbf{y}_1, \ldots, \mathbf{y}_k, \mathbf{x}_1, \ldots, \mathbf{x}_k\right) \\ &= \prod_{i=1}^{k} \prod_{t=1}^{n} \text{dnorm}\left(y_{it}, \beta_{0i} + \beta_{1i} x_{it}, \sigma_i^2\right) \\ &\times \prod_{i=1}^{k} \text{dmvnorm}\left(\boldsymbol{\beta}_i, \boldsymbol{\theta}, \Sigma\right) \times \prod_{i=1}^{k} \text{dgamma}\left(\frac{1}{\sigma_i^2}, \frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right), \end{aligned} \tag{5}$$

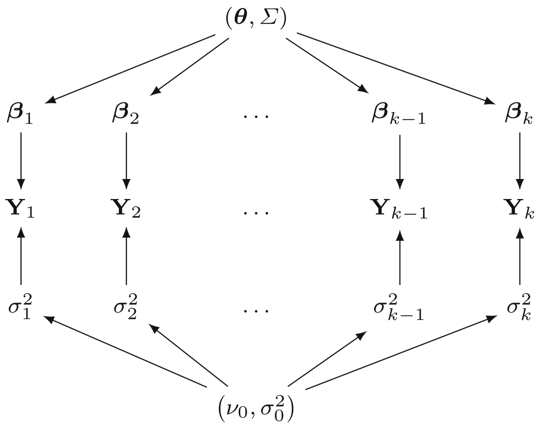


Fig. 2 Graphical representation of the hierarchical model

where `dnorm`, `dmvnorm`, and `dgamma` are the density functions for the *normal*, *multivariate normal*, and *gamma* distributions, respectively. Please note that the likelihood function provides information contained in the data for the unknown parameters of interests.

3.2 The prior distributions

In addition to the information contained in the data, some extra knowledge may come from experimenter's prior experience of the system giving rise to the data. Incorporation of this prior knowledge about the system in model building may increase the precision of the estimates for the parameters of interest.

The prior distribution of a parameter represents experimenter's prior belief via hyper-parameters. We note that the prior belief should be unbiased: a belief which reflects the expected truth of the system that generates the data. In addition, we emphasize that the prior belief should not be strong unless there is enough evidence towards the belief. This is the case because the strong prior belief might pull the posterior belief towards itself. In situations where there is weak or no prior belief, we suggest to be objective and follow the lead by the data.

The prior distribution for the course-specific error variance σ_0^2 is considered as

$$\sigma_0^2 \sim \text{Gamma}\,(\alpha_1, \alpha_2), \tag{6}$$

where the prior belief regarding σ_0^2 is represented by the hyper-parameters α_1 (the shape parameter) and α_2 (the rate parameter). Specifically, the expected prior belief about σ_0^2 is expressed as $E(\sigma_0^2) = \alpha_1/\alpha_2$. Here, small and large numbers of α_1 represent weak and strong prior belief regarding the course-specific error variance σ_0^2.

We restrict ν_0 to be a whole number and choose the prior on ν_0 to be a discrete analogue of exponential distribution on $\{1, 2, 3, \ldots\}$ as following:

$$p(\nu_0) \propto (1 - e^{-\alpha_3})e^{-\alpha_3 \nu_0}, \tag{7}$$

where α_3 reflects the strength of prior belief about ν_0. Specifically, small values of α_3 represent weak belief about ν_0 and vice versa.

The parameter vector for the course-specific mean $\boldsymbol{\theta}$ is considered to follow the following distribution

$$\boldsymbol{\theta} \sim \text{Multivariate-Normal}\,(\boldsymbol{\theta}_0, \Sigma_0)\,. \tag{8}$$

The prior belief about the course-specific mean vector is considered to be $\boldsymbol{\theta_0}$ (i.e., $E(\boldsymbol{\theta}) = \boldsymbol{\theta}_0$) and the strength of the prior belief is represented by the variance-covariance matrix Σ_0. Here, Σ_0 is a positive-definite matrix where the diagonal elements contain the variances (large values of the variances represent weak prior belief) and the off-diagonal elements contain the covariances (small values of absolute covariance represent weak correlation between θ's).

The prior distribution corresponding to the variance-covariance matrix Σ of the course-specific mean vector $\boldsymbol{\theta}$ is

$$\Sigma \sim \text{Inverse-Wishart}\left(\eta_0, S_0^{-1}\right), \tag{9}$$

where the prior belief about Σ is represented by S_0 (i.e., $E(\Sigma) = S_0/(\eta_0 - d - 1)$, where d be the dimension of $\boldsymbol{\theta}$) and the strength of prior belief is represented by η_0. As in other prior distributions, large and small values of η_0 represent strong and weak prior belief, respectively.

3.3 The posterior distributions

After collecting data, we combine the information from the data with prior belief to obtain posterior belief. In other words, the posterior belief is our updated belief after observing the data.

The posterior distribution for the overall error variance σ_0^2 is Gamma which is obtained using Eqs. (4) and (6) as following:

$$\sigma_0^2|\sigma_1^2, \ldots, \sigma_k^2, \nu_0 \sim \text{Gamma}\left(\alpha_1 + \frac{k\nu_0}{2}, \alpha_2 + \frac{\nu_0}{2}\sum_{i=1}^{k}\frac{1}{\sigma_i^2}\right). \tag{10}$$

The posterior distribution for the strength parameter ν_0 for the overall error variance σ_0^2 is obtained using Eqs. (4) and (7):

$$\nu_0|\sigma_0^2, \sigma_1^2, \ldots, \sigma_k^2 \propto \left(\frac{\left(\frac{\nu_0\sigma_0^2}{2}\right)^{\frac{\nu_0}{2}}}{\Gamma(\frac{\nu_0}{2})}\right)^k \times \left(\prod \frac{1}{\sigma_i^2}\right)^{\frac{\nu_0}{2}-1} e^{-\nu_0\left[\alpha_3 + \frac{\sigma_0^2}{2}\sum \frac{1}{\sigma_i^2}\right]} \tag{11}$$

The posterior distribution for the student-specific error variance $1/\sigma_i^2$ is independent Gamma which is obtained by using Eqs. (2) and (4):

$$\frac{1}{\sigma_i^2}|\mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\beta}_i, \nu_0, \sigma_0^2 \sim \text{Gamma}\left(\frac{\nu_0 + d}{2}, \frac{\nu_0\sigma_0^2 + \text{SSE}(\boldsymbol{\beta}_i)}{2}\right), \tag{12}$$

where $\text{SSE}(\boldsymbol{\beta}_i) = \left(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta}_i\right)^T\left(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta}_i\right)$ be the student-specific sum of squares of errors.

The posterior distribution for the student-specific mean vector $\boldsymbol{\beta}_i$ is independent Multivariate-Normal which is obtained using Eqs. (2) and (3) as following:

$$\boldsymbol{\beta}_i|\mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\theta}, \sigma_i, \Sigma \sim \text{Multivariate-Normal}\left(E\left(\boldsymbol{\beta}_i\right)_p, \text{Var-Cov}\left(\boldsymbol{\beta}_i\right)_p\right), \tag{13}$$

with posterior mean

$$E\left(\boldsymbol{\beta}_i\right)_p = \left(\frac{\mathbf{x}_i^T\mathbf{x}_i}{\sigma_i^2} + \Sigma^{-1}\right)^{-1}\left(\frac{\mathbf{x}_i^T\mathbf{y}_i}{\sigma_i^2} + \Sigma^{-1}\boldsymbol{\theta}\right)$$

and posterior variance-covariance matrix

$$\text{Var-Cov}\left(\boldsymbol{\beta}_i\right)_p = \left(\frac{\mathbf{x}_i^T \mathbf{x}_i}{\sigma_i^2} + \Sigma^{-1}\right)^{-1}.$$

The posterior distribution for the overall course-specific mean vector $\boldsymbol{\theta}$ is Multivariate Normal which is obtained using Eqs. (3) and (8):

$$\boldsymbol{\theta}|\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k, \Sigma \sim \text{Multivariate-Normal}\left(E\left(\boldsymbol{\theta}\right)_p, \text{Var-Cov}\left(\boldsymbol{\theta}\right)_p\right), \tag{14}$$

with posterior mean

$$E\left(\boldsymbol{\theta}\right)_p = \left(\Sigma_0^{-1} + k\Sigma^{-1}\right)^{-1}\left(\Sigma_0^{-1}\boldsymbol{\theta}_0 + \Sigma^{-1}\sum_{i=1}^{k}\boldsymbol{\beta}_i\right)$$

and posterior variance-covariance matrix

$$\text{Var-Cov}\left(\boldsymbol{\theta}\right)_p = \left(\Sigma_0^{-1} + k\Sigma^{-1}\right)^{-1}.$$

The posterior distribution of Σ is Inverse-Wishart which is obtained via Eqs. (3) and (9) as following

$$\Sigma|\boldsymbol{\theta}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k \sim \text{Inverse-Wishart}\left(\nu_0 + k, \left(S_{\boldsymbol{\theta}} + S_0\right)^{-1}\right), \tag{15}$$

where $S_{\boldsymbol{\theta}} = \sum_{i=1}^{k}\left(\boldsymbol{\beta}_i - \boldsymbol{\theta}\right)\left(\boldsymbol{\beta}_i - \boldsymbol{\theta}\right)^T$.

3.4 Distribution of missing values

Let $Y_{i,o}$ and $Y_{i,m}$ be the marks that are *observed* and *missing*, respectively, for the ith student in a course. Given the distribution of the marks

$$\left\{\mathbf{Y}_{i,o}, \mathbf{Y}_{i,m}|\mathbf{x}_i, \boldsymbol{\beta}_i, \Sigma, \sigma_i^2\right\} \sim \text{Multivariate-Normal}\left(\boldsymbol{\mu}_i = \mathbf{x}_i\boldsymbol{\beta}_i, V_i = \mathbf{x}_i\Sigma\mathbf{x}_i^T + \sigma_i^2 I_{d\times d}\right), \tag{16}$$

the missing marks for the ith student are imputed *independently* by generating data from the following distribution

$$\left\{Y_{i,m}|y_{i,o}, \boldsymbol{\beta}_i, \Sigma, \sigma_i^2\right\} \sim \text{Multivariate-Normal}\left(\boldsymbol{\mu}_{i,m|i,o}, V_{i,m|i,o}\right), \tag{17}$$

where

$$\boldsymbol{\mu}_{i,m|i,o} = \boldsymbol{\mu}_{i,m} + V_{i,m,o}\left(V_{i,o,o}\right)^{-1}\left(y_{i,o} - \boldsymbol{\mu}_{i,o}\right)$$

and

$$V_{i,m|i,o} = V_{i,m,m} - V_{i,m,o}\left(V_{i,o,o}\right)^{-1}V_{i,o,m}$$

are obtained using the properties of *conditional* multivariate normal distribution with

$$\boldsymbol{\mu}_i = \begin{bmatrix}\boldsymbol{\mu}_{i,o}\\ \boldsymbol{\mu}_{i,m}\end{bmatrix}$$

and

$$V_i = \begin{bmatrix}V_{i,o,o} & V_{i,o,m}\\ V_{i,m,o} & V_{i,m,m}\end{bmatrix}.$$

Note that, in generating the missing values, the $\boldsymbol{\beta}_i$, Σ and σ_i^2 are generated first from their respective posterior distributions. The computational details of the missing value imputation is provided in Sect. 3.5 with specifics in step 7 of the Gibbs sampling algorithm.

Many missing value imputation methods exist in the literature such as: mean imputation, row average imputation, ordinary least squares imputation, linear model based imputation, local least squares imputation [7], regression imputation, imputation of longitudinal data [30], singular value decomposition, principal component analysis [18], and expectation maximization [25]. Most of the above missing value imputation methods are classical methods which allow the imputation of missing values only once given the observed data. Our method of missing value imputation is fully Bayesian which allows seamless imputation of missing values before every generation of the MCMC scans of the parameters from their posterior distributions.

3.5 The Gibbs sampling algorithm

The approximation of posterior distribution via Gibbs sampling is briefly presented below. For a given state of the parameters

$$\left\{\nu_0^{(s)}, \sigma_0^{2(s)}, \boldsymbol{\beta}_1^{(s)}, \ldots, \boldsymbol{\beta}_k^{(s)}, \sigma_1^{2(s)}, \ldots, \sigma_k^{2(s)}, \boldsymbol{\theta}^{(s)}, \Sigma^{(s)}, \mathbf{y}_m^{(s)}\right\}$$

and $\mathbf{y}_i = \{\mathbf{y}_{i,o}, \mathbf{y}_{i,m}^{(s)}\}$, a new state is generated as follows:

1. Sample ν_0 using

$$\nu_0^{(s+1)} \sim P\left(\nu_0|\sigma_0^{2(s)}, \sigma_1^{2(s)}, \ldots, \sigma_k^{(2)}\right),$$

where the posterior distribution is specified in Eq. (11).

2. Sample σ_0^2 using

$$\sigma_0^{2(s+1)} \sim P\left(\sigma_0^2|\sigma_1^{2(s)}, \ldots, \sigma_k^{2(s)}, \nu_0^{(s+1)}\right),$$

with posterior distribution specified in Eq. (10).

3. For each $i \in \{1, 2, \ldots, k\}$, independently sample σ_i^2 using

$$\sigma_i^{2(s+1)} \sim P\left(\sigma_i^2|\mathbf{y}_{i,o}, \mathbf{y}_{i,m}^{(s)}, \mathbf{x}_i, \boldsymbol{\beta}_i^{(s)}, \nu_0^{(s+1)}, \sigma_0^{2(s+1)}\right),$$

where the posterior distribution is specified in Eq. (12).

4. For each $i \in \{1, 2, \ldots, k\}$, independently sample $\boldsymbol{\beta}_i$ using

$$\boldsymbol{\beta}_i^{(s+1)} \sim P\left(\boldsymbol{\beta}_i|\mathbf{y}_{i,o}, \mathbf{y}_{i,m}^{(s)}, \mathbf{x}_i, \boldsymbol{\theta}^{(s)}, \sigma_i^{2(s+1)}, \Sigma^{(s)}\right),$$

with posterior distribution specified in Eq. (13).

5. Sample $\boldsymbol{\theta}$ using

$$\boldsymbol{\theta}^{(s+1)} \sim P\left(\boldsymbol{\theta}|\boldsymbol{\beta}_1^{(s+1)}, \ldots, \boldsymbol{\beta}_k^{(s+1)}, \Sigma^{(s)}\right),$$

where the posterior distribution is specified in Eq. (14).

6. Sample Σ using

$$\Sigma^{(s+1)} \sim P\left(\Sigma|\boldsymbol{\theta}^{(s+1)}, \boldsymbol{\beta}_1^{(s+1)}, \ldots, \boldsymbol{\beta}_k^{(s+1)}\right),$$

with posterior distribution specified in Eq. (15).

7. For each $i \in \{1, 2, \ldots, k\}$, independently sample the missing marks using

$$\mathbf{y}_{i,m}^{(s+1)} \sim P\left(\mathbf{Y}_{i,m}|\mathbf{y}_{i,o}, \mathbf{y}_{i,m}^{(s)}, \mathbf{x}_i, \boldsymbol{\beta}_i^{(s+1)}, \Sigma^{(s+1)}, \sigma_i^{2(s+1)}\right),$$

where the posterior distribution of missing data given the observed data is specified in Eq. (17).

The order in which the new parameters and missing data are generated does not matter. What does matter is that each parameter is updated conditional upon the current value of the remaining parameters and imputed missing values. The Gibbs sampling algorithm is implemented using the R [22] language.

3.6 Specification of hyperparameters

The hyper-parameters for the prior distributions are chosen as described below. The hyper-parameters for the prior distribution of overall error variance σ_0^2 (Eq. (6)) are chosen as $\alpha_1 = 1$ and $\alpha_2 = 1/100$. This specification implies that $E(\sigma_0^2) = 100$. In other words, our prior specification considers large variability between $\sigma_1^{-2}, \sigma_2^{-2}, \ldots, \sigma_k^{-2}$. That is, our belief implies large overall error variability and small within student error variability. At the same time, we consider α_3 (the hyper-parameter for the prior of ν_0) to be 1 as the small values of α_3 represent weak prior belief about ν_0. Given that we have no prior knowledge about the overall error variance and the student specific error variances, a weak prior imposes less subjectivity and lets the data objectively determine the parameter values.

To choose the hyper-parameters for the prior distribution of $\boldsymbol{\theta}$ (Eq. (8)), we fitted ordinary least squares (OLS) regression to student-specific marks and saved the OLS coefficients. The OLS coefficients are averaged to specify $\boldsymbol{\theta}_0$ (the prior mean vector for $\boldsymbol{\theta}$). The prior variance-covariance matrix Σ_0 is considered to be the sample covariance of the OLS coefficients. Such a prior distribution represents belief that is aligned with the information contained in the data.

Similarly, we consider the prior sum of squares matrix S_0 (Eq. (6)) to be equal to the sample covariance of the ordinary least squares estimates of the coefficients. This specification ensures the lead by the data as the expectation of the prior distribution of Σ equals to the ordinary least squares estimates of the coefficients. But at the same time, we consider $\eta_0 = d + 2 = 4$. This specification makes the prior distribution flat or diffuse to make the prior belief weak. Such a prior specification ensures less subjectivity or more objectivity.

In order to initiate the Gibbs sampling algorithm, the initial parameter values are obtained from the OLS estimates of the coefficients. The initial missing values are obtained by simple average of the student-specific observed marks. We used the first 1000 scans of the Gibbs sampler as burn-in and threw the values of the realizations out. This eliminates the effect of the initial values we have chosen to start the Gibbs sampler. We then ran the Gibbs sampler for another 10,000 scans and saved every 10th scan to produce a sequence of 1000 values for each parameter. We checked if the Markov chain for each parameter obtained stationarity or not. The autocorrelation values and plots for the sequence of saved scans for each parameter are also examined. After confirming convergence in terms of stationarity and minimal autocorrelation, we proceeded next to perform Bayesian posterior inference.

Figure 3 shows the MCMC trace-plots against thinned scans of the chains for θ_0 (top-left) and θ_1 (top-right). These plots show the generated values of θ_0 and θ_1 from their respective posterior distributions saved at every 10th scan of the chain after throwing out the burned-in scans. It is visible that the chain has achieved stationarity. The bottom panels of Fig. 3 show

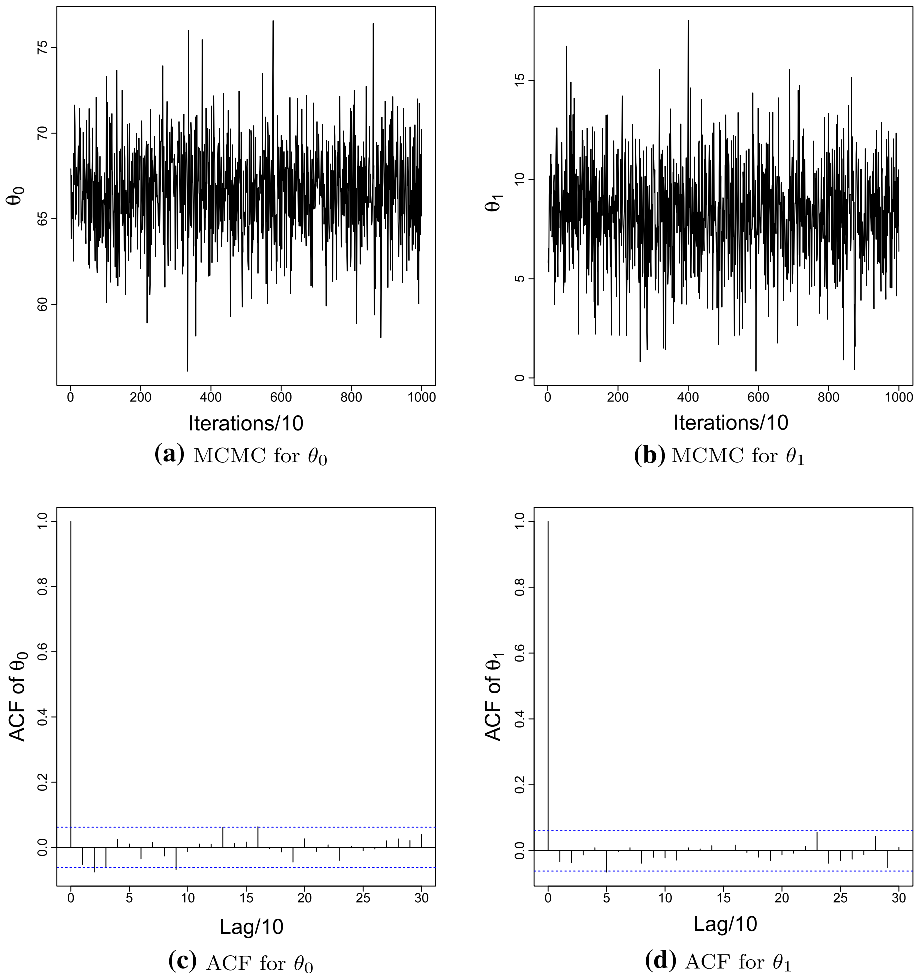


Fig. 3 Trace plots for MCMC samples (top) and autocorrelation functions (bottom) for θ_0 and θ_1 for MATH 1070: Mathematics for Business and Economics

the auto-correlation functions (ACFs) for θ_0 (bottom-left) and θ_1 (bottom-right). It is also clear that the generated thinned scans of the chains are nearly uncorrelated. In fact, their effective sample sizes are straight 1000 in each. Similarly, we have confirmed stationarity for the parameters for every other course selected in our study.

4 Results

We summarized the number of students, the percentage of disengaged students, and the percentage of missing marks within each course in Table 1. The percentage of disengaged students is defined as the percentage of students within a course who did not participate in any of the assessment/evaluation components after classes and assessments/evaluations transitioned to online. The percentage of disengaged students varies from 0% to 5.56%

Table 2 Number of evaluation components and percentage of hands-on (lab, coding and programming) components

Courses	Number of evaluation components	Hands-on percentages
MATH1070	7	0
MATH1240	12	0
MATH1250	10	0
MATH1640	12	0
MATH2200	11	0
MATH2240	9	0
STAT2000	9	0
ARET1400	15	26.7
ARET2600	20	30
COMP2680	15	60
COMP4980	12	50

across different courses. Even though these numbers are relatively small, the implications are significant. We investigated the reasons as to why students disengaged after March 15, 2020. According to students' records/responses, at least 43% of them were in need of special supports and accommodations due to mental illness caused by concussion, severe disability in coping with stress, and being slow in processing information. Most of such students had difficulties following the course content in an online delivery mode without face-to-face interactions with the instructors. The other 57% of disengaged students were those who struggled the most with the course contents. Since there were no marks available for the evaluation components for the disengaged students after March 15, 2020, we excluded them from further analysis.

The percentage of missing marks is defined as the percentage of missing evaluation components relative to all evaluation components for the remaining students in a course. These numbers vary from 3.73 to 14.07%. As shown in Table 2, the percentage of missing marks are higher for the courses in which more direct hands-on supports were required in the form of labs, coding, programming, and seminars than the courses evaluated mainly using assignments, quizzes, tests, and exams.

4.1 Course-specific analysis

Figure 4 and Table 4 show the comparisons of overall performances (top-left panel) and three student-specific performances (top-right, bottom-left and bottom-right panels) before and after March 15 for the course MATH 1070: Mathematics for Business and Economics. The overall marks for all the students in this course went up from a median of 66.683% (95% credible interval of 61.203–71.763%) to 74.991% (with 95% credible interval of 69.353–80.647%). Even though there is a small overlap in the 95% credible intervals for the overall marks, student-specific marks for some students increased by a large margin. The first case (Student 5) made a significant shift of marks from a failing letter grade to a passing letter grade of *C* (TRU grading scales are summarized in Table 3). This student was most likely failing as his/her average marks had a 95% credible interval ranging from 47.384% to 50.470%. The second case (Student 14) made an even larger shift from a failing letter grade to a strong letter

Table 3 Grading scale for undergraduate academic programs at Thompson Rivers University

Letter grade	Numerical grade	Grade points
A+	90–100	4.33
A	85–89	4.00
A−	80–85	3.67
B+	77–79	3.33
B	73–76	3.00
B−	70–72	2.67
C+	65–69	2.33
C	60–64	2.00
C−	55–59	1.67
D	50–54	1.00
F	0–49	0.00

Table 4 Overall performances and interesting cases for MATH 1070: Mathematics for Business and Economics

Index	Statistics	COVID-19	
		Before	After
Overall	Lower limit	61.203	69.353
	Median	66.683	74.991
	Upper limit	71.763	80.647
	SD	2.682	2.845
Student 5	Lower limit	47.384	61.043
	Median	48.827	63.618
	Upper limit	50.470	66.098
	SD	0.764	1.252
Student 14	Lower limit	41.539	74.345
	Median	45.215	81.242
	Upper limit	50.226	86.742
	SD	2.213	3.109
Student 27	Lower limit	52.770	85.118
	Median	54.246	87.676
	Upper limit	55.663	89.824
	SD	0.735	1.140

grade of $A-$. The third case (Student 27) also showed a large shift from a barely passing letter grade of D to a passing letter grade of A. None of the distributions of marks before and after March 15 had any overlaps, making the shifts highly statistically significant.

Figures 5, 6, 7 and Tables 5, 6, 7 show the comparisons of overall performance and student-specific cases in each of the courses MATH 1240: Calculus 2, MATH 1250: Calculus for Biological Sciences 2, and MATH 1640: Technical Mathematics 1. After the transition to online delivery, the overall medians in these courses were increased by about 20%, 8% and 7%, respectively. The overall increase of marks was statistically significant for the course MATH 1240. On the other hand, the increase of student-specific marks within the course was not uniform. For example, the performance of Student 5 in MATH 1240 improved by only 1.67%, while Students 6 and 29 had an increase of marks by 33.66% and 29.18%,

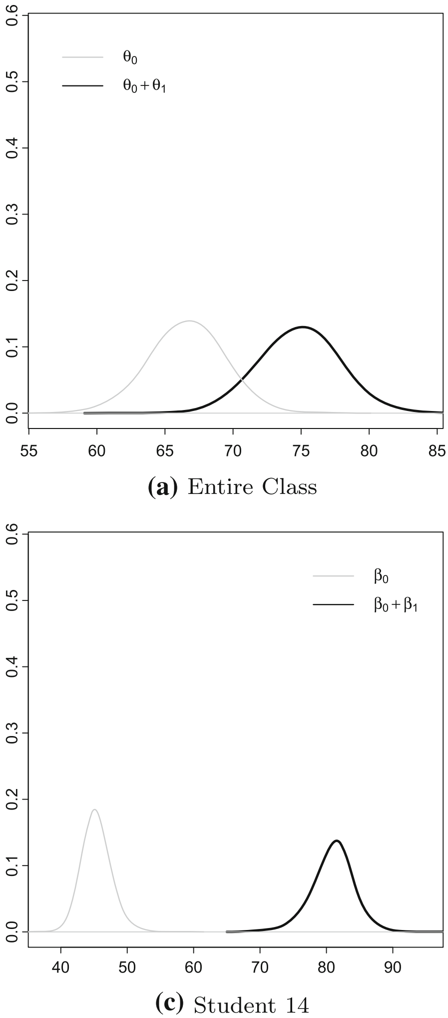


Fig. 4 Overall performances and interesting cases for MATH 1070: Mathematics for Business and Economics

respectively. The last two students shifted their marks from a failing grade to passing letter grades of A^- and B, respectively. A more or less similar trend is observed for Students 6 and 24 in MATH 1250, with a small variability in their grades. In other words, a few weaker students managed to improve their marks consistently to a significantly higher range of grades after March 15. However, some good-standing students, such as Student 26 in MATH 1250 and Student 5 in MATH 1640, experienced a decline in their marks. Note that the decline of marks for these good students were statistically insignificant. While the decline of marks for Student 5 in MATH 1640 was practically insignificant, the decline for Student 26 in MATH 1250 was considered practically significant as the change of the letter grade was from A^+ to A.

Figures 8, 9 and Tables 8, 9 show the comparisons of overall marks and student-specific cases for the second-year courses MATH 2200: Introduction to Analysis, and MATH 2240:

Table 5 Overall performances and interesting cases for MATH 1240: Calculus 2

Index	Statistics	COVID-19	
		Before	After
Overall	Lower limit	50.752	70.954
	Median	57.443	77.655
	Upper limit	63.931	84.020
	SD	3.256	3.317
Student 5	Lower limit	49.610	49.650
	Median	54.974	56.645
	Upper limit	60.289	64.299
	SD	2.683	3.730
Student 6	Lower limit	45.104	77.593
	Median	47.317	80.984
	Upper limit	49.515	84.194
	SD	1.143	1.658
Student 29	Lower limit	42.372	70.368
	Median	44.162	73.348
	Upper limit	45.991	76.251
	SD	0.919	1.488

Table 6 Overall performances and interesting cases for MATH 1250: Calculus for Biological Sciences 2

Index	Statistics	COVID-19	
		Before	After
Overall	Lower limit	58.007	67.038
	Median	65.298	73.057
	Upper limit	72.111	79.085
	SD	3.607	3.105
Student 6	Lower limit	51.292	77.433
	Median	53.754	80.809
	Upper limit	56.679	83.703
	SD	1.398	1.659
Student 24	Lower limit	46.274	62.440
	Median	48.279	65.002
	Upper limit	50.673	67.650
	SD	1.085	1.279
Student 26	Lower limit	87.516	84.943
	Median	90.833	88.016
	Upper limit	93.522	91.788
	SD	1.480	1.745

Table 7 Overall performances and interesting cases for MATH 1640: Technical Mathematics 1

Index	Statistics	COVID-19	
		Before	After
Overall	Lower limit	75.653	81.511
	Median	79.584	86.368
	Upper limit	83.197	91.297
	SD	1.993	2.454
Student 5	Lower limit	95.986	94.657
	Median	96.652	95.672
	Upper limit	97.313	96.739
	SD	0.350	0.506
Student 12	Lower limit	53.355	74.291
	Median	56.996	79.491
	Upper limit	60.992	84.669
	SD	1.941	2.654
Student 25	Lower limit	57.587	75.424
	Median	63.029	83.916
	Upper limit	70.621	91.720
	SD	3.217	3.985

Differential Equations 1, respectively. MATH 2200 is a proof course, required for Math Majors and a gateway to heavy-proof math courses. After the transition to online delivery, the overall median marks were increased by about 10%. For most of the students the improvement was not notable. However, for some students, such as Students 5 and 9, the improvements in marks (41.96% and 41.28%, respectively) were practically and statistically significant. Both of the students' marks moved from a failing letter grade to passing letter grades of B^+ and C^-, respectively. In MATH 2240, the overall marks increased by about 11% with almost zero overlap before and after COVID-19. A similar trend of growth in the marks, especially for the struggling students, has been observed in this course. For instance, Students 14 and 21 have shown a significant jump in their marks after March 15. On the other hand, the A^+ level Student 29 experienced a very minimum change in his/her grade. This change is insignificant both practically and statistically.

Figure 10 and Table 10 display the comparisons of marks before and after March 15 for the course STAT 2000: Probability and Statistics. The overall marks went up slightly from a median of 65.653–67.675%. As there is an overlap in the two distributions, the increase is not statistically significant. Regarding the student-specific cases, Students 5 and 22 who were barley passing the course had an increase by about 15% and 16%, respectively. None of the two distributions overlap each other before and after March 15, hence, the shifts are highly statistically significant. On the other hand, a few good-standing students experienced a decline in their marks from higher percentages to lower percentages. For example, Student 12 was in the A^+ grade range before March 15, while his/her marks went down significantly by around 11% to the letter grade of $B+$ after March 15.

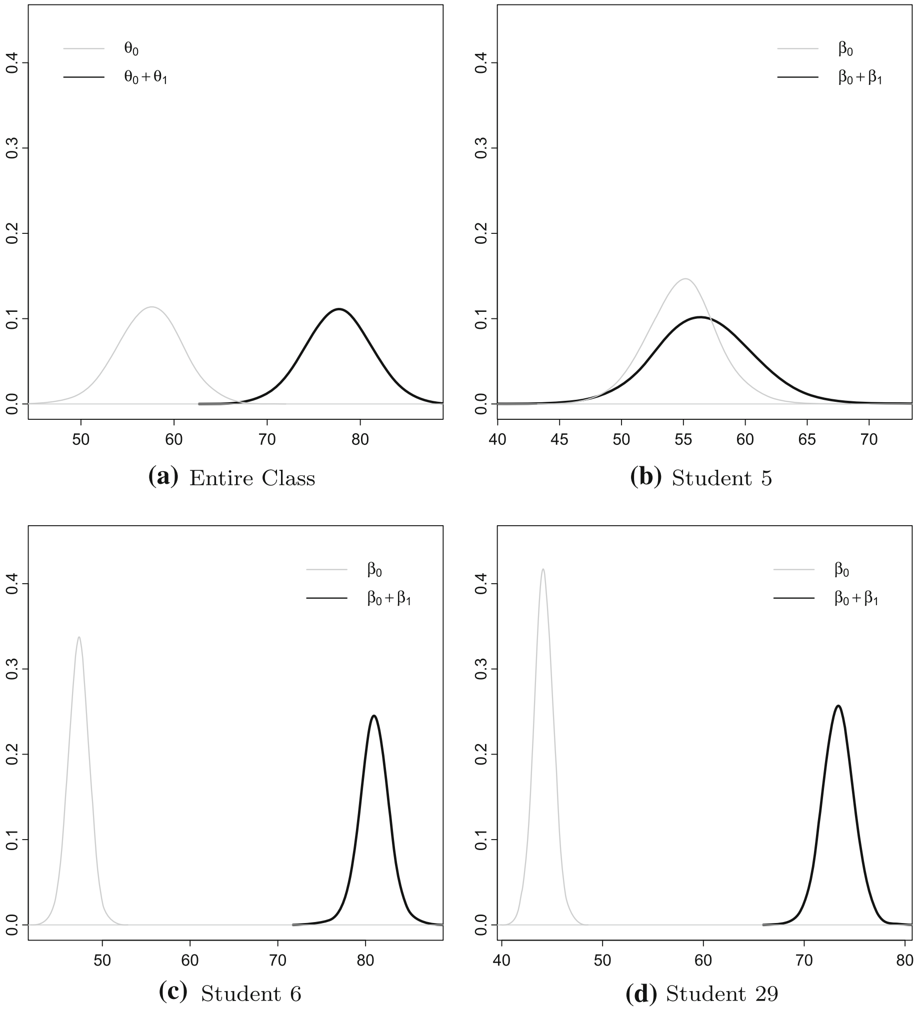
Figures 11, 12 and Tables 11, 12 compare the overall performance and student-specific cases for the courses ARET 1400: Civil Technology 1, and ARET 2600: Statics and Strength of Materials. The overall students' performances in these two courses declined after the transition to online delivery. In ARET 1400, the median marks decreased significantly by

Table 8 Overall performances and interesting cases for MATH 2200: Introduction to Analysis

Index	Statistics	COVID-19	
		Before	After
Overall	Lower limit	50.782	64.223
	Median	61.399	71.676
	Upper limit	71.497	79.387
	SD	5.266	3.993
Student 5	Lower limit	31.422	70.392
	Median	36.763	78.721
	Upper limit	43.312	85.561
	SD	2.933	3.996
Student 9	Lower limit	14.921	55.265
	Median	16.164	57.443
	Upper limit	17.428	59.480
	SD	0.633	1.043
Student 17	Lower limit	83.927	83.610
	Median	89.224	90.670
	Upper limit	94.369	97.751
	SD	2.646	3.661

Table 9 Overall performances and interesting cases for MATH 2240: Differential Equations 1

Index	Statistics	COVID-19	
		Before	After
Overall	Lower limit	60.021	71.390
	Median	65.683	76.214
	Upper limit	71.397	81.067
	SD	2.909	2.461
Student 14	Lower limit	58.863	77.163
	Median	59.293	77.717
	Upper limit	59.773	78.204
	SD	0.225	0.252
Student 21	Lower limit	39.808	68.417
	Median	43.660	73.472
	Upper limit	49.035	77.973
	SD	2.270	2.450
Student 29	Lower limit	93.213	93.281
	Median	94.972	94.961
	Upper limit	96.569	96.903
	SD	0.818	0.904



(a) Entire Class (b) Student 5 (c) Student 6 (d) Student 29

Fig. 5 Overall performances and interesting cases for MATH 1240: Calculus 2

12.14%, whereas in ARET 2600, the median marks decreased by 4.93%. The decrease in ARET 2600 is not statistically significant. The percentage of decrease or increase of marks varies from one student to another. For example, the decrease in marks for Student 29 in ARET 1400 and Student 15 in ARET 2600 were 30.75% and 18.74%, respectively. On the other hand, Student 6 in ARET 1400 and Student 4 in ARET 2600 did not experience a large drop in their marks. For Student 4 in ARET 2600, the decrease is insignificant both statistically and practically. There are a few exceptions to this trend as well. For example, Student 12 in ARET 1400 and Student 12 in ARET 2600 did experience some increase in their marks after transitioning to online delivery mode. Again, the increase of 2.36% for Student 12 in ARET 1400 was insignificant both statistically and practically.

Figures 13, 14 and Tables 13, 14 show overall performances and student-specific cases for the courses COMP 2680: Web Development, and COMP 4980: Bioinformatics. Students'

Table 10 Overall performances and interesting cases for STAT 2000: Probability and Statistics

Index	Statistics	COVID-19	
		Before	After
Overall	Lower limit	59.315	62.279
	Median	65.653	67.675
	Upper limit	72.045	73.220
	SD	3.226	2.802
Student 5	Lower limit	51.056	65.219
	Median	51.849	66.706
	Upper limit	52.668	68.160
	SD	0.420	0.747
Student 12	Lower limit	89.326	77.278
	Median	90.761	79.874
	Upper limit	91.950	82.403
	SD	0.666	1.234
Student 22	Lower limit	52.412	77.009
	Median	53.504	79.135
	Upper limit	54.587	80.874
	SD	0.571	0.991

Table 11 Overall performances and interesting cases for ARET 1400: Civil Technology 1

Index	Statistics	COVID-19	
		Before	After
Overall	Lower limit	82.387	67.281
	Median	85.131	72.933
	Upper limit	87.859	79.046
	SD	1.411	2.907
Student 6	Lower limit	91.180	86.049
	Median	92.351	88.559
	Upper limit	93.529	90.766
	SD	0.595	1.175
Student 12	Lower limit	84.014	85.621
	Median	84.843	87.205
	Upper limit	85.634	88.729
	SD	0.426	0.813
Student 29	Lower limit	84.563	51.266
	Median	86.614	55.863
	Upper limit	88.573	60.593
	SD	1.019	2.408

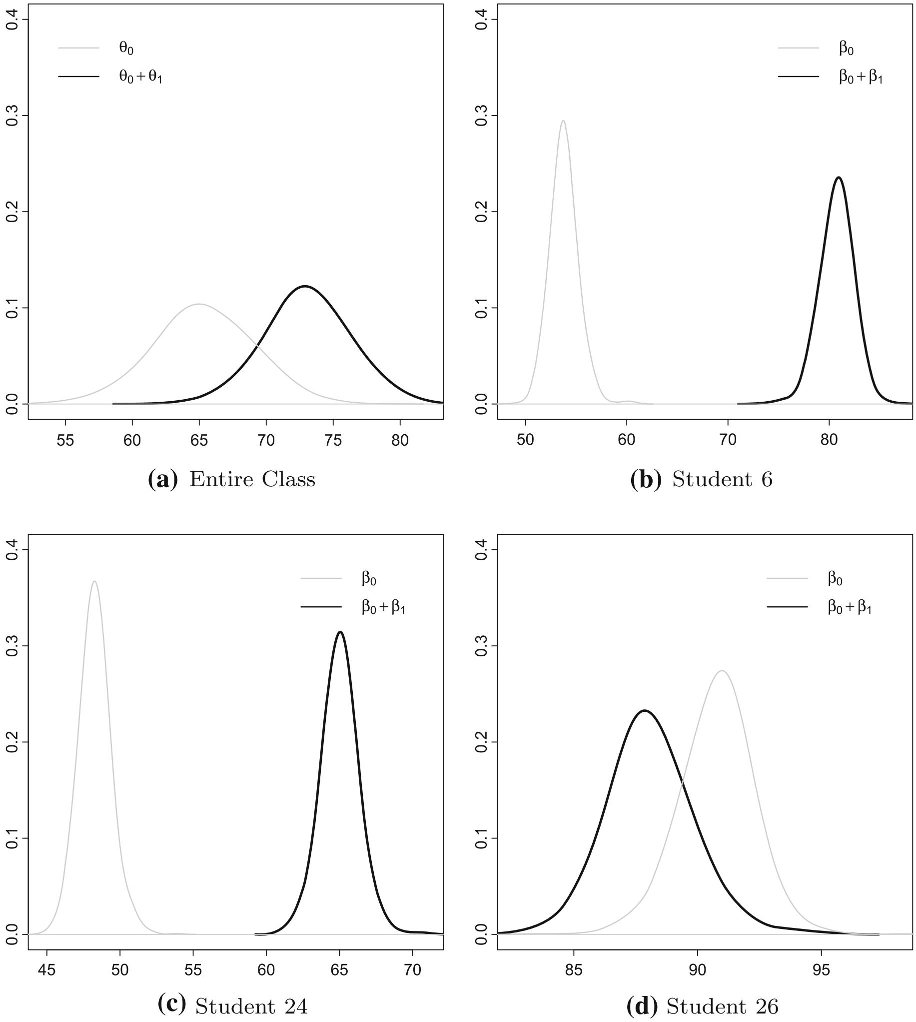


Fig. 6 Overall performances and interesting cases for MATH 1250: Calculus for Biological Sciences 2

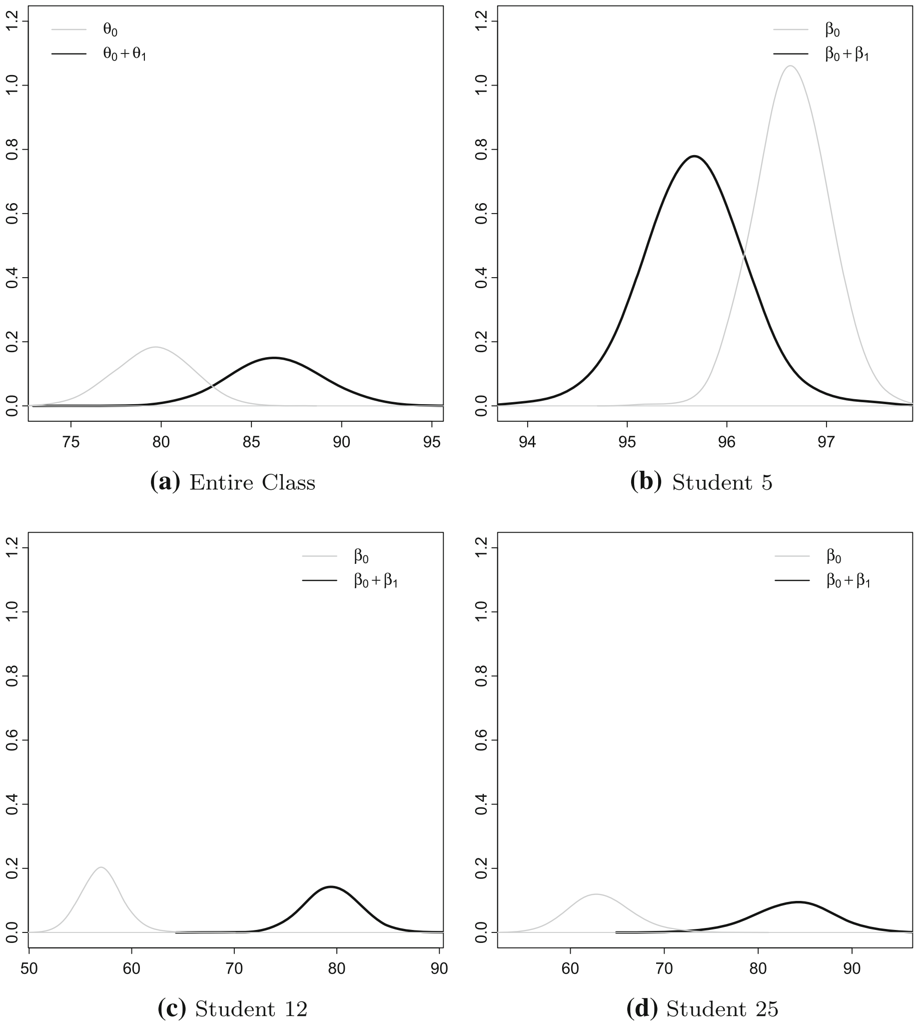
performances in these Computing Science courses were negatively affected by the transition to online delivery mode. The median of overall marks went down by 3% for COMP 2680 and by 19% for COMP 4980. The decrease is statistically insignificant for COMP 2680, while significant for COMP 4980. The marks decreased for a large number of students after March 15. For example, the performance of Students 18 and 22 in COMP 2680 decreased by 14.86% and 33.82% and the performance of Students 8 and 11 in COMP 4980 decreased by 1.95% and 40.56%, respectively. On the other hand, some good-standing students were able to maintain their good performance after March 15, such as Student 16 in COMP 2680 and Student 19 in COMP 4980. However, these increases were insignificant both practically and statistically.

Table 12 Overall performances and interesting cases for ARET 2600: Statics and Strength of Materials

Index	Statistics	COVID-19	
		Before	After
Overall	Lower limit	74.795	67.592
	Median	78.120	73.199
	Upper limit	80.904	78.190
	SD	1.580	2.755
Student 4	Lower limit	78.975	76.232
	Median	79.537	77.069
	Upper limit	80.089	77.871
	SD	0.278	0.423
Student 12	Lower limit	72.386	86.270
	Median	73.467	88.037
	Upper limit	74.548	89.737
	SD	0.558	0.881
Student 15	Lower limit	73.177	52.305
	Median	74.522	55.779
	Upper limit	75.826	59.157
	SD	0.659	1.750

Table 13 Overall performances and interesting cases for COMP 2680: Web Site Design and Development

Index	Statistics	COVID-19	
		Before	After
Overall	Lower limit	89.318	84.949
	Median	91.591	88.953
	Upper limit	93.654	92.900
	SD	1.086	2.021
Student 16	Lower limit	93.964	94.873
	Median	94.950	95.962
	Upper limit	95.988	97.036
	SD	0.510	0.552
Student 18	Lower limit	97.425	82.524
	Median	97.869	83.006
	Upper limit	98.249	83.491
	SD	0.208	0.236
Student 22	Lower limit	90.997	56.612
	Median	92.492	58.664
	Upper limit	94.024	60.570
	SD	0.758	1.019



(a) Entire Class **(b)** Student 5 **(c)** Student 12 **(d)** Student 25

Fig. 7 Overall performances and interesting cases for MATH 1640: Technical Mathematics 1

5 Discussion of results

Among the 11 courses in this study, both increasing and decreasing trends in students' marks were observed. Specifically, a general increase in the marks is shown in theory-based courses requiring lower-level cognitive skills according to Bloom's Taxonomy of Knowledge, whereas a general decrease in the marks are shown in the courses requiring either interactive hands-on support or higher-level cognitive skills.

5.1 Rising trend in courses requiring lower-level cognitive skills

University-level math courses are normally delivered in a traditional lecture format with the instructor teaching core concepts and theories accompanied by related examples and

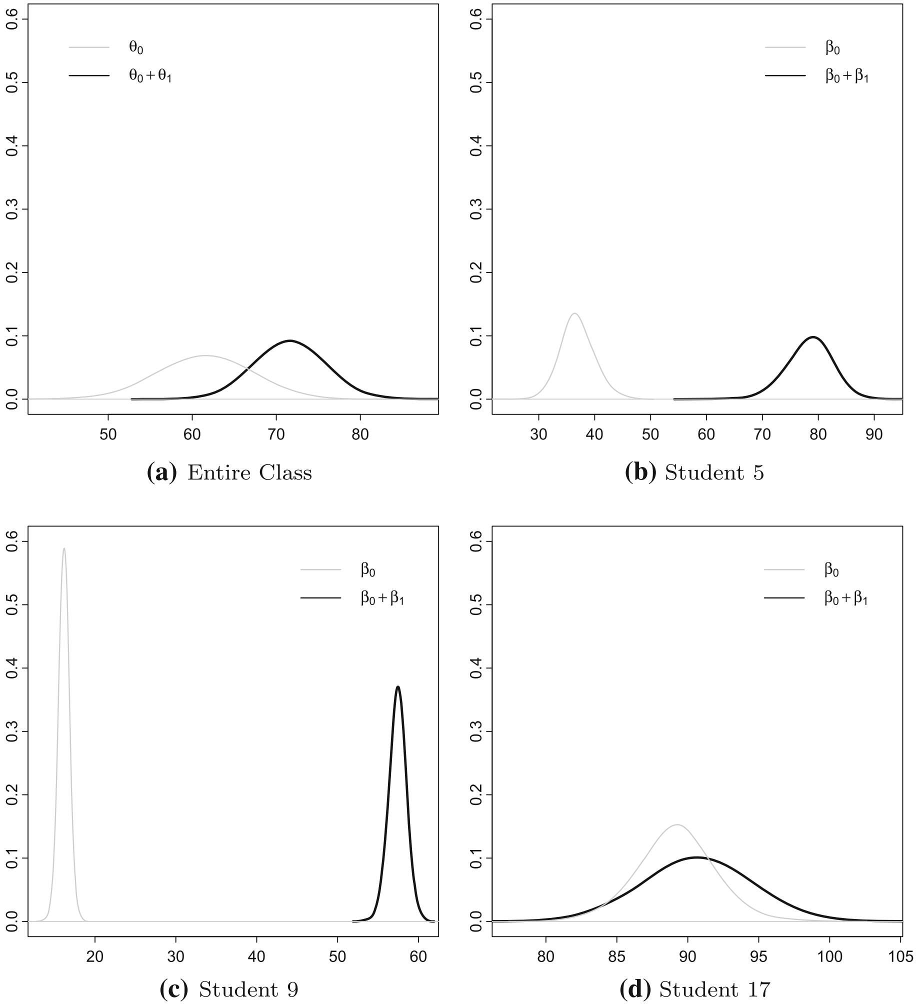
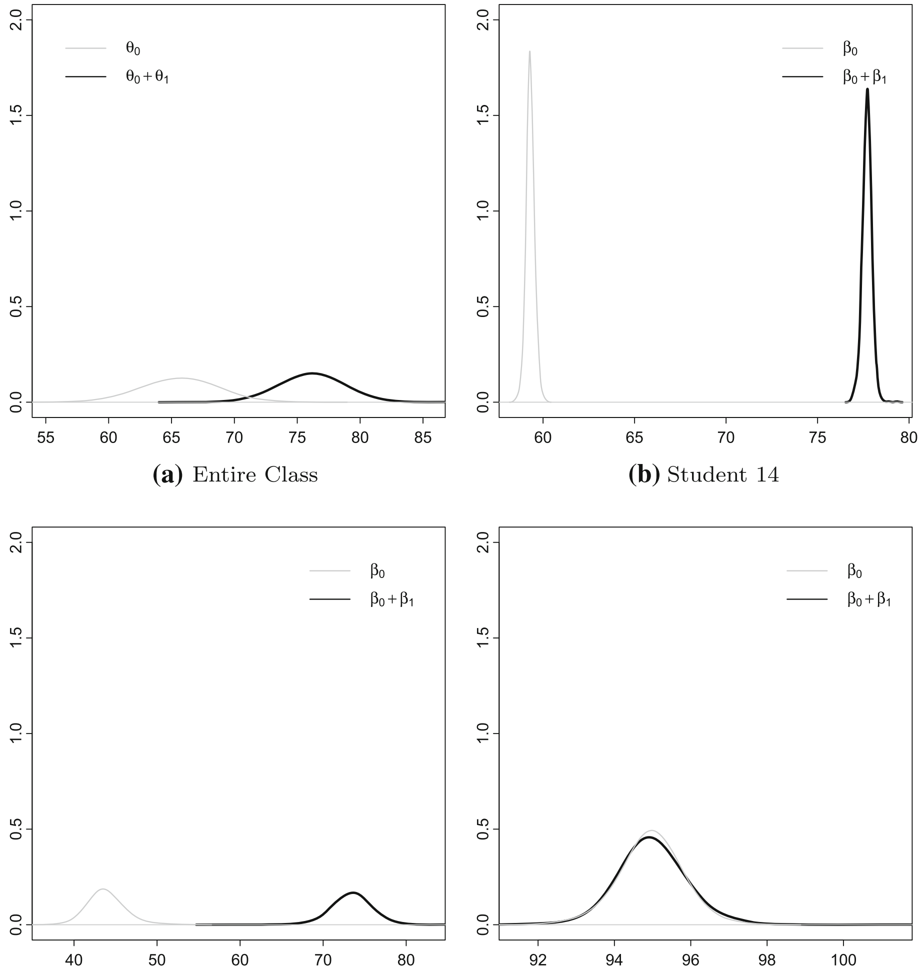


Fig. 8 Overall performances and interesting cases for MATH 2200: Introduction to Analysis

applications varying from direct applications to more conceptual and intricate ones. Student assessment is then composed of several in-class quizzes, written homeworks, one or two midterms and a final exam with generally more weight on summative assessments than formative ones. Knowing that a standard first- or second-year math course is often taken by a large group of students enrolled in various university programs with a wide range of backgrounds in math, assessments in these courses are mainly focused on questions with a medium-difficulty level in order to reasonably evaluate students' learning. In other words, according to Bloom's Taxonomy of Knowledge, in-person math exams are normally testing low- to medium-level skills and abilities with limited allocation of questions to higher level skills such as analysis and synthesis. However, transitioning to online and open-book exams did change all equations.

The unprecedented closure of universities due to COVID-19 imposed an unexpected shock to academia in particular to the traditional culture of course delivery and assessment in



(a) Entire Class **(b)** Student 14

(c) Student 21 **(d)** Student 29

Fig. 9 Overall performances and interesting cases for MATH 2240: Differential Equations 1

mathematics. Given the very limited time for instructors to prepare for switching to online modes of delivery, creating online open-book tests and restructuring in-person exams to be suitable for the online format was rather infeasible. Moreover, although some instructors did attempt to design tests relatively different than in-person exams in order to target deeper levels of understanding, they were faced with students' complaints and resistance. This is completely understandable though on account of the lack of training throughout the semester for such exams. Students in MATH 1250, for example, even in normal circumstances, are mostly categorized in the struggling group with levels of math anxiety; hence it is unrealistic to expect them to perform well in an exam format to which they are not used to. Therefore, most MATH and STAT courses maintained a format similar to in-person exams with the online assessments being open-book.

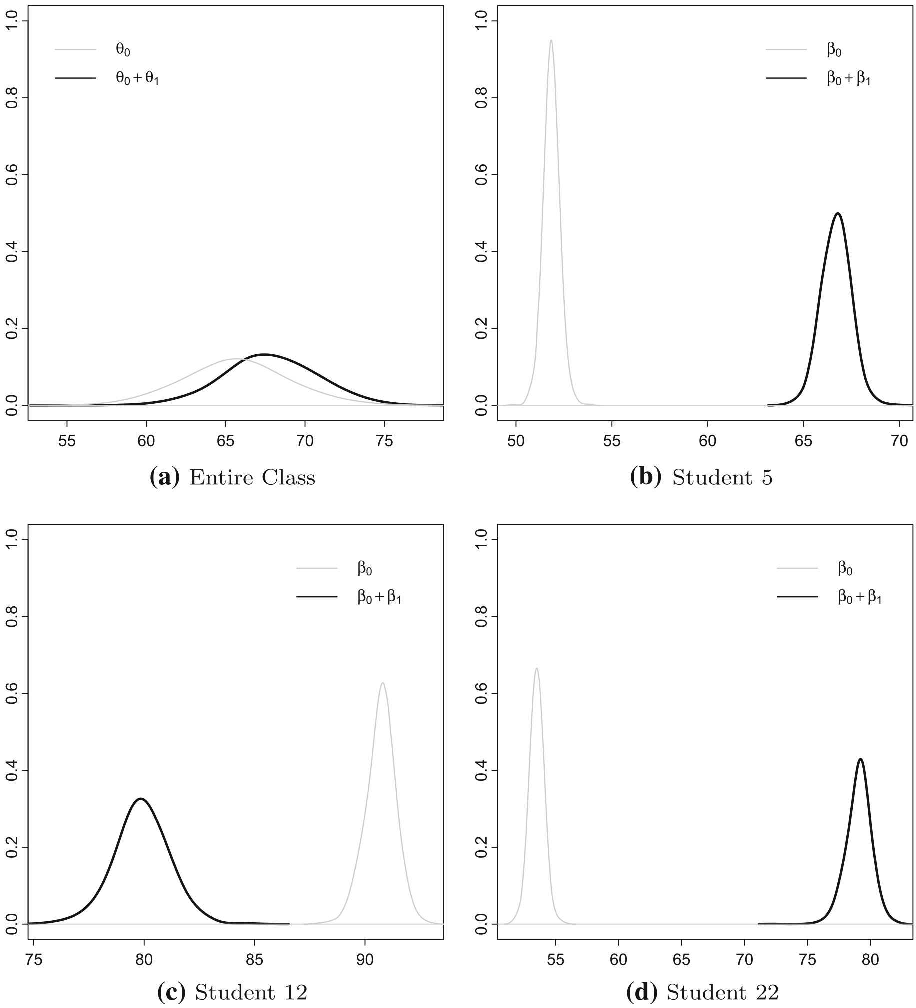


Fig. 10 Overall performances and interesting cases for STAT 2000: Probability and Statistics

Nevertheless, with the availability of resources in an open-book exam the low- and medium-level question types that target memory and comprehension skills such as recalling, defining, describing or explaining concepts were no longer truly examining students' learning as these can be easily found in textbooks, class notes, etc. It was observed that students had a better performance in these question types in the online version of exams compared to similar face-to-face exams prior to university closure. For instance, students in MATH 1240 are normally in the category of moderately strong Science students and their performance improved after the transition to online delivery. The improvement took place mostly because the assessment components and their structures did not change, except the exams being open-book, and students could adapt to the new delivery method with relative ease. Similarly, in MATH 2240 which is a second-year course, the overall performance has grown significantly which can be partly attributed to open-book exams maintaining a similar structure to face-to-face ones.

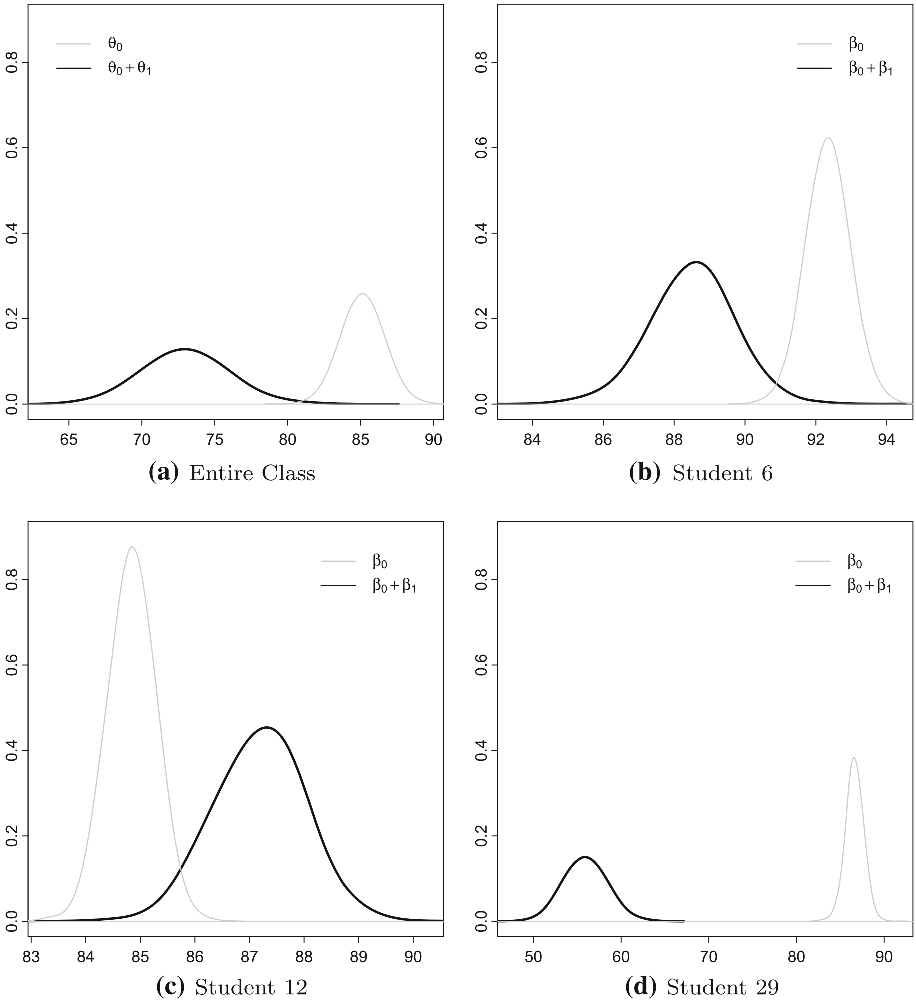


Fig. 11 Overall performances and interesting cases for ARET 1400: Civil Technology 1

Furthermore, one can consider the role of technology and online resources available to students during an online math exam. Tools were no longer limited to a basic scientific calculator. Advanced online calculators and math programs, along with many online forums were at hand during an open-book exam. For example, at the beginning of the Winter 2020 semester, all students in MATH 2200 were struggling with the course content as writing a rigorous proof is a skill never taught in the first-year courses. After about a month, many students in this class could improve their proof writing skills and consequently improved their grades to some extent. When the course transitioned to online delivery, the assessment components remained unchanged, most of the students continued their upward trend and their overall performance improved. But this transition might have provided some weaker students the opportunity to seek other resources for answers in non-invigilated tests. This pattern can be seen for students 5 and 9 (Fig. 8). This is also observed in Student 24 in MATH 1250 and Student 21 in MATH 2240, whose large grade improvements turned their grade

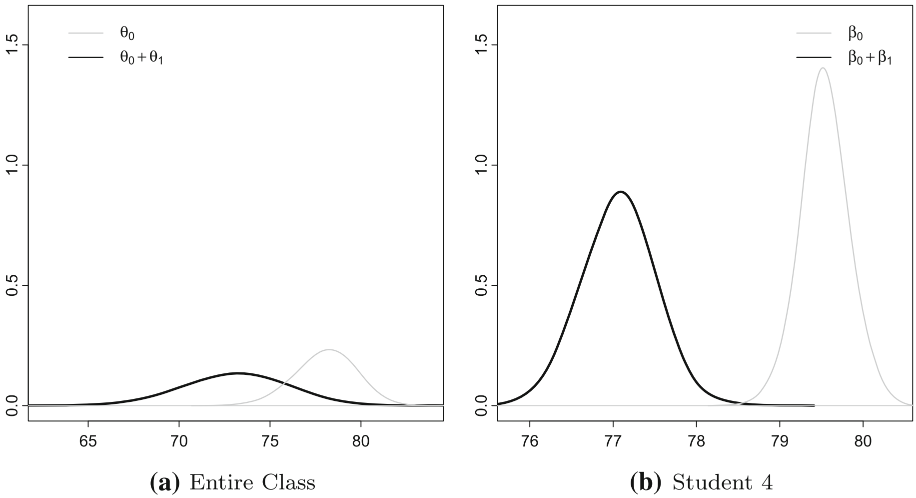


Fig. 12 Overall performances and interesting cases for ARET 2600: Statics and Strength of Materials

from failing to the letter grades of C^+ and B, respectively. The performance of Student 29 in MATH 2240 and Student 17 in MATH 2200, however, did not change significantly after March 15. These students were among the average- to high-performing students.

The statistics course (STAT 2000) we investigated in this study tests medium-level cognitive skills of students, as per Bloom's Taxonomy, and falls slightly above the Math courses. This course requires students to understand the methods, organize the information in the data, apply the methodology to the data to gain insightful knowledge, and provide summary and explanations of the findings. Nurturing these skills requires some discussions between the instructor and the students. After the transition to online delivery, the support that students needed were provided to the best of the instructor's ability, especially to the students who requested support via online meetings and discussions. As the structure of the course is very similar to Math courses with some added applications, the overall performance of students

Table 14 Overall performances and interesting cases for COMP 4980: Introduction to Bioinformatics

Index	Statistics	COVID-19	
		Before	After
Overall	Lower limit	80.916	58.211
	Median	83.927	64.378
	Upper limit	86.774	70.547
	SD	1.494	3.133
Student 8	Lower limit	97.972	95.892
	Median	98.401	96.444
	Upper limit	98.815	97.033
	SD	0.208	0.284
Student 11	Lower limit	77.107	35.682
	Median	81.164	40.602
	Upper limit	85.171	45.960
	SD	2.059	2.690
Student 19	Lower limit	88.281	88.216
	Median	90.469	91.361
	Upper limit	92.440	94.331
	SD	1.048	1.482

improved slightly by about 2%. This increase of overall performance is not as large as the Math courses which are aligned with lower-level cognitive skills. Moreover, when most of the weaker students in this course improved their grades by a large margin, the top students experienced a slight decrease in their marks despite their potential.

5.2 Decreasing trend in courses requiring higher-level cognitive skills and interactive hands-on support

In ARET courses, application and analysis of learned theories and concepts play a key role in the assessments. While assessments in math courses revolve around defining, calculating or reproducing facts pertaining to a topic, ARET courses demand students' expertise in applying the knowledge learned to novel applications/situations. Assessments for the ARET courses investigated in this study entail developing problem solving skills with emphasis on analyzing and devising the concepts and principles in applied situations. These skills are categorized as medium- to high-level cognitive skills, as per Bloom's Taxonomy. Students in these ARET courses experienced a decline in their grades after March 15 mainly because they needed some hands-on and face-to-face support to develop their analyzing and problem solving skills required for the final exam. Providing such hands-on support was not feasible after March 15 and, as a result, many students in ARET 1400 and ARET 2600 did not perform well in their final exam. A few of these students happened to be among the stronger cohort of students: for example, Student 29 in ARET 1400 and Student 15 in ARET 2600.

Computing Science is known as a 'learning-by-doing' subject, and most COMP courses require enormous interactive supports in hands-on programming and laboratories, without which it is hard for students to succeed in these courses. Table 2 shows that the computing science courses at TRU are more geared towards hands-on practices in lab components compared to MATH and STAT courses. For the two COMP courses investigated in this

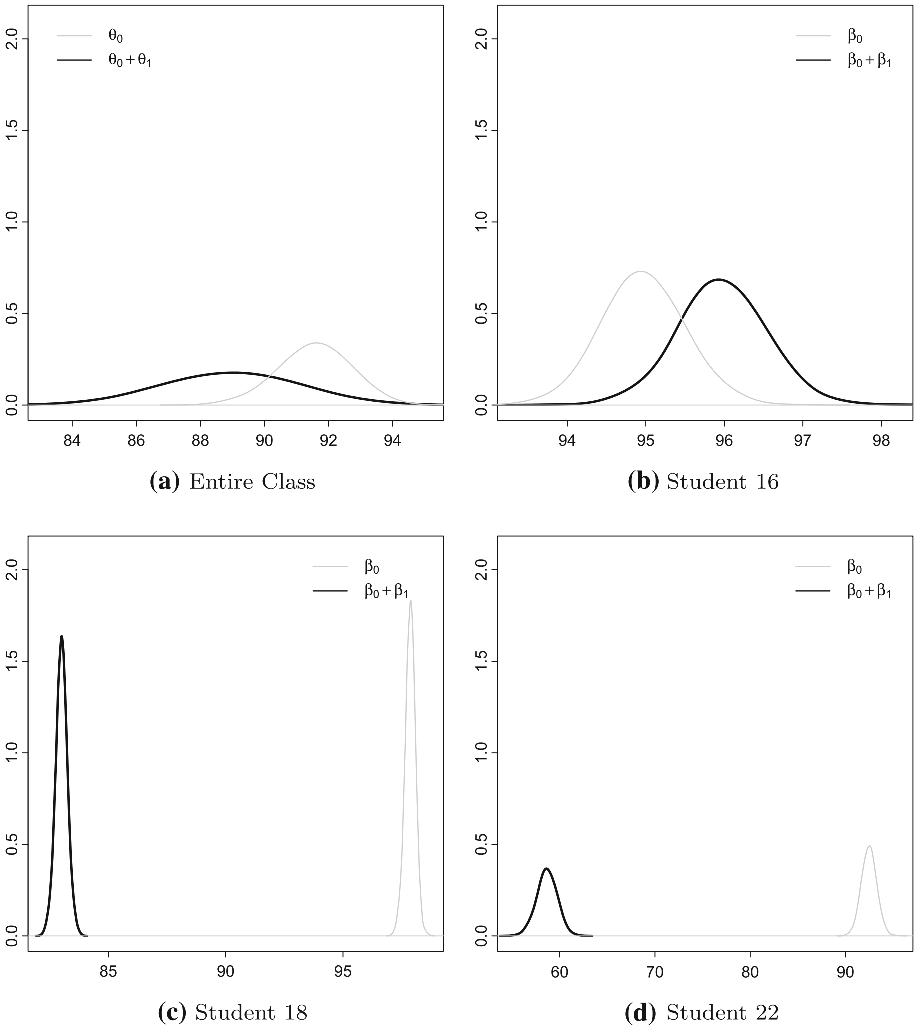


Fig. 13 Overall performances and interesting cases for COMP 2680: Web Site Design and Development

study, students in COMP 2680 learn web development skills in lectures, then practice and apply these skills in hands-on laboratories. COMP 4980 is an interdisciplinary course where students learn how to apply computing science skills to analyze, synthesize and interpret the biological data. It requires not only hands-on laboratories to practice problem solving skills, but also supports from instructor's domain-specific knowledge to help students connect the biological problems that they try to solve with computational models and interpret their results both biologically and mathematically. In other words, COMP courses investigated in this study test students' medium- to high-level cognitive skills, as per Bloom's Taxonomy.

Unsurprisingly, students' performances in both courses were negatively affected by the rapid switch from face-to-face to online delivery mode. The marks dropped after March 15 for most students as shown in Figs. 13 and 14, including students with good standings (e.g., Student 18 in COMP 2680 and Student 8 in COMP 4980) and relatively weak students (e.g.,

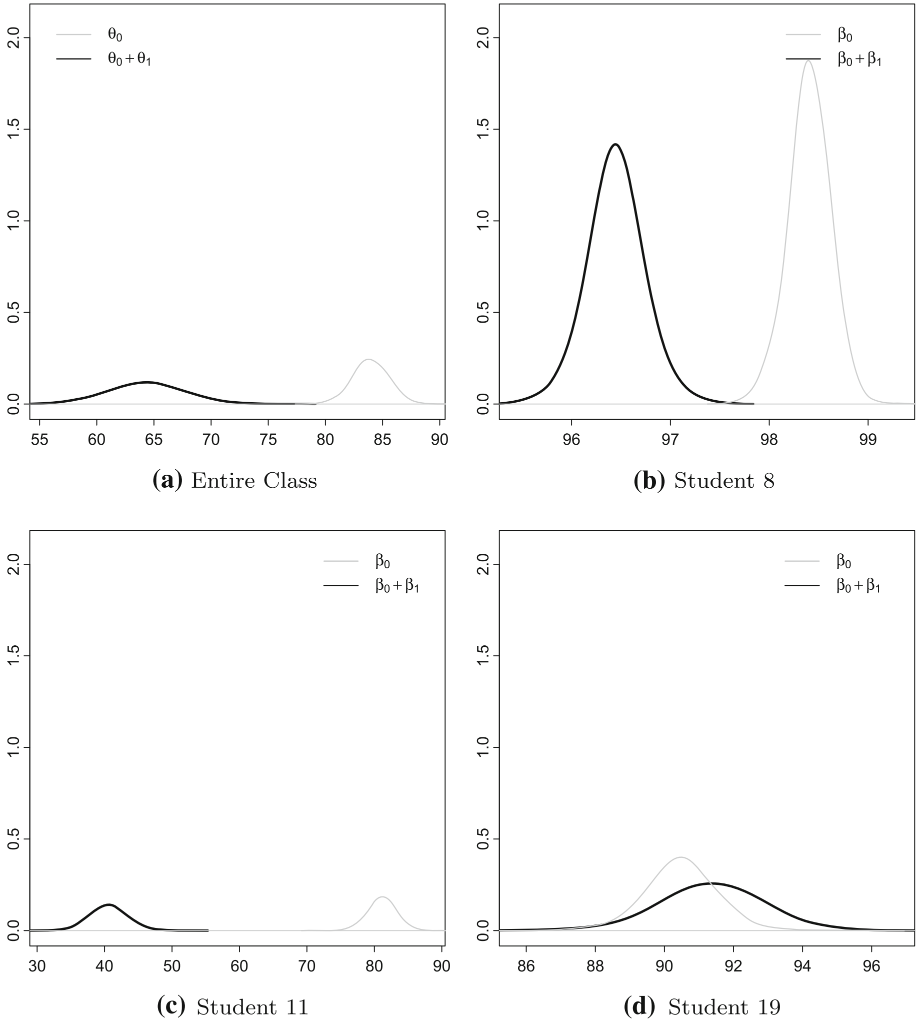


Fig. 14 Overall performances and interesting cases for COMP 4980: Introduction to Bioinformatics

Student 22 in COMP 2680 and Student 11 in COMP 4980). There might be a few reasons that potentially caused the decrease of students' marks after March 15 in these courses. For instance, most students did better in the first half of the semester, because the first few topics of the course were introductory topics which were easy to pick up, and more difficult topics were introduced in the second half of the semester (after March 15). During the face-to-face delivery, students could also get hands-on help from instructors or TAs in lectures, labs, or Computing Science Help Center, but the hands-on labs and the Help Center were no longer accessible after March 15. For COMP 4980, there was a group project in the last three weeks of the semester, but students could not get face-to-face interaction with each other and did not receive the same level of support from the instructor due to the online delivery mode, therefore, some students struggled in the term project.

6 Conclusion

We report results from a moderately large scale study from 11 courses, where the effects of COVID-19 on students' performance were compared with empirical rigor. This study shows that a sudden change of delivery mode has an immense impact on students' marks. After switching to online delivery mode and assessments due to COVID-19, students' marks were increased in theory-based courses that required lower-level cognitive skills based on Bloom's Taxonomy, whereas in courses with hands-on lab, coding and programming components, or courses that required higher-level cognitive skills, the marks were decreased. The larger increase (for MATH and STAT courses) or decrease (for COMP and ARET courses) of marks are mainly observed for weaker students as opposed to stronger students. The group of stronger students experienced a smaller decrease of marks, while some very hard working students were able to maintain a good standing of marks towards their credentials. The impact has been much more significant on students with special needs who disengaged from the course after March 15. We also emphasize on the fact that the COVID-19 outbreak, lock-down and closure of schools have exposed students to an extraordinary stress level. Students faced the sudden shock of online transition with virtually no education and training on how to take ownership over their submitted work in the online space and be accountable for that.

The authors of this paper observed similar trends in results across Canada which was discussed in many educational workshops and meetings, such as SSC Webinar on Teaching Statistics Online[4] and CMS COVID-19 Research and Education Meeting (CCREM).[5] Hence, the results of the study can also be generalized across Canada. This is because most, if not all, of the universities across Canada follow the same educational system and experienced moving towards online teaching and assessment at around the same time.

Our novel contribution is analysing and comparing COVID-19 effects on students' marks. In addition to this novel application, we also designed and developed novel computational models. A Bayesian linear mixed effect model was designed to fully address the comparison of marks. The implementation of Bayesian missing value imputation is novel both in terms of statistics and application.

In this paper, we considered a normal distribution (Eq. (2)) for the response variable of interest. As alternatives, one may wish to use other probability distributions as they fit. For example, in the presence of unusually small or large numbers in student-specific data, one may wish to use heavy-tailed non-central t-distribution. The use of alternative distributions may complicate the computational process for posterior realizations in situations when the posterior distributions are not in closed form. In such a situation, one may need to use Metropolis or Metropolis-Hastings algorithm instead of Gibbs sampling. On the other hand, the applications of open-source MCMC software, such as JAGS [10], WinBUGS [20], or Stan [6] may appear handy to improve computational issues.

This paper considered STEM courses offered under the faculty of science at TRU. On the other hand, consideration of more courses across multiple faculties in University might be of interest. Such interest may also extend to multiple universities in a country or across the world. However, such augmentation of data may require one to use the multilevel linear mixed effects model [21].

Acknowledgements We thank Mr. Shi Shi for sharing the course information of COMP 2680, and Dr. Musfiq Rahman for his discussions earlier in this study. We also thank the Editor, and two anonymous reviewers for their insightful comments, which substantially strengthened the arguments presented here.

4 https://ssc.ca/en/node/10244.

5 https://ccrem20.cms.math.ca/.

References

1. Arias, J.J., Swinton, J., Anderson, K.: Online vs. face-to-face: a comparison of student outcomes with random assignment. e-J. Bus. Educ. Scholarsh. Teach. **12**(2), 23 (2018). https://files.eric.ed.gov/fulltext/EJ1193426.pdf
2. Ashby, J., Sadera, W.A., McNary, S.W.: Comparing student success between developmental math courses offered online, blended, and face-to-face. J. Interact. Online Learn. **10**(3), 10 (2011). https://www.ncolr.org/jiol/issues/pdf/10.3.2.pdf
3. Basilaia, G., Kvavadze, D.: Transition to online education in schools during a SARS-CoV-2 Coronavirus (COVID-19) pandemic in Georgia. Pedag. Res. **5**, 1–9 (2020). https://doi.org/10.29333/pr/7937
4. Bhaskaran, K., Smeeth, L.: What is the difference between missing completely at random and missing at random? Int. J. Epidemiol. **43**(4), 1336–1339 (2014). https://doi.org/10.1093/ije/dyu080
5. Cao, W., Fang, Z., Hou, G., Han, M., Xu, X., Dong, J., Zheng, J.: The psychological impact of the covid-19 epidemic on college students in china. Psychiatry Res. **287**(112934), (2020). https://doi.org/10.1016/j.psychres.2020.112934
6. Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A.: Stan: a probabilistic programming language. J. Stat. Softw. **76**(1), 1–32 (2017). https://doi.org/10.18637/jss.v076.i01. https://www.jstatsoft.org/v076/i01
7. Celton, M., Malpertuy, A., Lelandais, G., Brevern, A.G.d.: Comparative analysis of missing value imputation methods to improve clustering and interpretation of microarray experiments. BMC Genomics **11**(15) (2010). https://doi.org/10.1186/1471-2164-11-15
8. Cronbach, L.J.: Coefficient alpha and the internal structure of tests. Psychometrika **16**, 297–334 (1951). https://doi.org/10.1007/BF02310555
9. Dennon, A.: Coronavirus impacts on students and online learning (2020). https://www.bestcolleges.com/blog/coronavirus-impacts-on-students/
10. Denwood, M.J.: runjags: an R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. J. Stat. Softw. **71**(9), 1–25 (2016). https://doi.org/10.18637/jss.v071.i09
11. Doreleyers, A., Knighton, T.: COVID-19 pandemic: academic impacts on postsecondary students in canada. Statistics Canada (2020). https://www150.statcan.gc.ca/n1/pub/45-28-0001/2020001/article/00015-eng.htm
12. Geyer, C.J.: Practical Markov chain Monte Carlo. Stat. Sci. **7**(4), 473–483 (1992). https://www.jstor.org/stable/2246094
13. Jones, S.J., Vena, M.L.: Learning equity between online and on-site mathematics courses. MERLOT J. Online Learn. Teach. **9**(1), 1–12 (2013). https://jolt.merlot.org/vol9no1/jones_0313.pdf
14. Kleiner, C., Lord, M.: The cheating game: 'everyone's doing it,' from grade school to graduate school. U.S. News World Rep. **127**, 54–63 (1999)
15. Laird, N., Lange, N., Stram, D.: Maximum likelihood computations with repeated measures: application of the EM algorithm. J. Am. Stat. Assoc. **82**(397), 97–105 (1987). https://doi.org/10.2307/2289129. http://www.jstor.org/stable/2289129
16. Laird, N.M., Ware, J.H.: Random-effects models for longitudinal data. Biometrics **38**(4), 963–974 (1982). https://doi.org/10.2307/2529876. http://www.jstor.org/stable/2529876
17. Lawrence, C., Altschul, S., Boguski, M., Liu, J., Neuwald, A., Wootton, J.: Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. Science **262**(5131), 208–214 (1993). https://doi.org/10.1126/science.8211139. https://science.sciencemag.org/content/262/5131/208
18. Lin, W., Tsai, C.F.: Missing value imputation: a review and analysis of the literature (2006–2017). Artif. Intell. Rev. **53**, 1487–1509 (2019)
19. Lindstrom, M.J., Bates, D.M.: Newton–Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. J. Am. Stat. Assoc. **83**(404), 1014–1022 (1988). https://doi.org/10.2307/2290128. http://www.jstor.org/stable/2290128
20. Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D.: Winbugs—a bayesian modelling framework: concepts, structure, and extensibility. Stat. Comput. **10**, 325–337 (2000). https://doi.org/10.1023/A:1008929526011
21. Pinheiro, J.C., Chao, E.C.: Efficient Laplacian and adaptive gaussian quadrature algorithms for multilevel generalized linear mixed models. J. Comput. Graph. Stat. **15**(1), 58–81 (2006). https://doi.org/10.1198/106186006X96962
22. R Core Team: R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna (2020). https://www.R-project.org
23. Rey, J.G.: The effects of online courses for student success in basic skills mathematics classes at california community colleges. Ph.D. thesis, University of Southern California (2010). https://search.proquest.com/openview/8872a71d22481e7919cdb8d71eda5992/1?pq-origsite=gscholar&cbl=18750&diss=y

24. Sahu, P.: Closure of universities due to coronavirus disease 2019 (COVID-19): impact on education and mental health of students and academic staff. Cureus **12(4)**, e7541 (2020). https://doi.org/10.7759/cureus.7541. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7198094/
25. Schafer, J.L., Yucel, R.M.: Computational strategies for multivariate linear mixed-effects models with missing values. J. Comput. Graph. Stat. **11**(2), 437–457 (2002). https://doi.org/10.1198/106186002760180608
26. Sintema, E.: Effect of COVID-19 on the performance of grade 12 students: implications for STEM education. EURASIA J. Math. Sci. Technol. Educ. **16**(7), em1851 (2020). https://doi.org/10.29333/ejmste/7893
27. Weems, G.H.: Comparison of beginning algebra taught onsite versus online. J. Dev. Educ. **26**(1), 10 (2002). https://www.learntechlib.org/p/95604/
28. Williams, J.B., Wong, A.: The efficacy of final examinations: a comparative study of closed-book, invigilated exams and open-book, open-web exams. Br. J. Educ. Technol. **40**, 227–236 (2009). https://doi.org/10.1111/j.1467-8535.2008.00929.x
29. Wong, E.T.: Impact of E-learning during SARS outbreak in Hong Kong. The social and cognitive impacts of e-commerce on modern organizations, pp. 185–203 (2004). https://doi.org/10.4018/978-1-59140-249-7.ch009
30. Zhang, Z.: Missing data imputation: focusing on single imputation. Ann. Transl. Med. **4**(1) (2016). http://atm.amegroups.com/article/view/8839
31. Zhao, Y., Staudenmayer, J., Coull, B.A., Wand, M.P.: General design bayesian generalized linear mixed models. Stat. Sci. **21**(1), 35–51 (2006). https://doi.org/10.2307/27645735. http://www.jstor.org/stable/27645735
32. Zu, Z.Y., Jiang, M.D., Xu, P.P., Chen, W., Ni, Q.Q., Lu, G.M., Zhang, L.J.: Coronavirus disease 2019 (COVID-19): a perspective from china. Radiology **0**(0), 200490 (2020). https://doi.org/10.1148/radiol.2020200490