A review of mathematical modelling in educational research in Indonesia

L Vitoria¹,², Marwan Ramli³, R Johar⁴ and M Mawarpury⁵

¹ Graduate School of Mathematics and Applied Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia
² Department of Elementary Education, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia
³ Department of Mathematics, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia
⁴ Department of Mathematics Education, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia
⁵ Department of Psychology, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia

E-mail: marwan.math@unsyiah.ac.id.

Abstract. Mathematical models are mathematical representations of real-life problems which are used to simplify the problems and also for future prediction. Mathematical modelling has been used in educational research around the world mainly to predict students’ academic performance and to identify essential factors that affect students’ learning. However, in Indonesia there are very few published researches that employ mathematical modelling for educational studies. The present study is an integrative review conducted to document these studies. The objective is to provide an overview of the usage of mathematical modelling in educational research in Indonesia. The method of the review follows Torraco’s guidelines of integrative review which consist of critical analysis and synthesis. Five open-access published articles were discovered. These articles were written either in English or in Bahasa Indonesia. 3 out of 5 articles focused on explaining several predictor variables that influence student performance. The other 2 articles focused on comparing different methods of modelling to predict student performance. These findings show that mathematical modelling has not been extensively utilized in educational research in Indonesia.

1. Introduction
Mathematical modelling is a cyclical procedure of producing a mathematical model that best fits a certain real-life problem so that the problem can be solved mathematically [1]. Using a mathematical model, the robustness and complexity of a problem can be presented in the form of a mathematical language [2]. Representing a problem in the form of a mathematical model provides several benefits: the preciseness of mathematical language can facilitate the process of formulating and identifying of fundamental ideas; the conciseness of mathematical language provides clear and well-defined rules for mathematical procedures; and mathematical language uses numbers and operators that can be calculated using computers [2]. Mathematical modelling has been employed for several reasons, for example to identify relationships between attributes of a problem or a system, to examine the outcome of a manipulation in a system, and to help with decision making [2]. For these reasons, mathematical modelling has been used in various research fields such as economy, biology, health and education.
In educational research, mathematical modelling has been employed to predict student performance [3] and to identify essential factors that affect student learning [4]. Unfortunately, in Indonesia, there are very few published researches that employ mathematical modelling as a method for educational studies. Most of the articles that are available online in Indonesian context focus on teaching mathematical modelling to students, they do not use mathematical model to predict student performance or to explain student academic achievement. Therefore, the present study intends to locate and review published researches that utilize mathematical modelling for these purposes. No similar study has been conducted for Indonesian context. Thus, this review is the first publication to provide this information.

As stated earlier, the present study intends to give an overview on the usage of mathematical modelling in educational research in Indonesia. The motivation to conduct this study was driven by the fact that the quality of Indonesian education compared to other countries is relatively low, based on global education attainment assessments such as Programme for International Student Assessment (PISA) and World University Rankings (https://www.timeshighereducation.com/world-university-rankings). The implementation of mathematical modelling in educational research may be useful to help identify factors that affect Indonesian students’ academic achievement and improve the quality of teaching and learning in Indonesian schools and universities. By conducting this review, we intend to map published studies that have been conducted in this field and to find gaps to help pave the way for future research in Indonesian educational research. Therefore, the following research questions were addressed: 1) what were the objectives of using mathematical modelling in the published researches? 2) what mathematical modelling methods were used in the published researches? 3) what are the prospects for future research in Indonesia regarding mathematical modelling?

2. Method

The present study utilized an integrative review method. According to Torraco [5], an integrative review is “a form of research that reviews, critiques, and synthesizes representative literature on a topic in an integrated way such that new frameworks and perspectives on the topic are generated”. The method of the review follows Torraco’s guidelines of integrative review which consist of critical analysis and synthesis. In critical analysis stage, main ideas, research methods, and key findings of the published researches were examined and the key contributions of the researches were identified along with their limitations [5]. In the synthesis stage, new information such as new agenda for future research, was generated based on the critical analysis of the published researches [5].

The criteria of the inclusion of the published researches were as follow: the works employed mathematical modelling techniques in the context of Indonesian educational studies; the works employed mathematical modelling as a research tool, not as a teaching strategy or teaching approach; the articles were written in either English or Bahasa Indonesia; the articles were published in the last 20 years that is from year 2000 until 2020; and the articles were open-access.

In order to locate published articles that fit the above criteria, a search strategy was proposed. The search strategy involved search engines or databases and search terms. The search engines used in the present review were the Garuda portal, which is a database of accredited Indonesian publications, and google search engine. The search terms used in Garuda were pemodelan matematika and mathematical modelling. The search terms used in google search engine were: student performance AND system OR application OR method OR process OR technique OR methodology OR procedure AND prediction OR estimation OR assessment AND mathematical modelling AND Indonesia.

The above search strategy returned numerous results, but only 5 articles met the criteria of inclusion for the present integrative review. The review stages including critical analysis and synthesis are presented in the following section.

3. Results and discussion

There were 5 published researches that satisfied the criteria of inclusion for the present integrative review. Table 1 displays the key attributes of the researches.
| Article no. | Year of publication | Authors | Language | Research Objectives | Methodology | Key findings |
|------------|---------------------|---------|----------|---------------------|-------------|--------------|
| 1          | 2019                | N Mardiana & A Faqih | Indonesian | To choose the best Partial Least Square Structural Equation Modelling (SEM-PLS) model to describe the relationship between LMS quality and students’ final score. | PLS-SEM algorithm | 1 alternative model was chosen as the best model based on Q2 criteria and other 6 criteria. |
| 2          | 2019                | F Alyani & Nurafni | English | To investigate the influence of gender and educational background on student performance. | PLS modelling | Gender, mid-test and final test influence GPA. |
| 3          | 2019                | Khasanah A U & Harwati | English | To estimate student performance using linear regression (LR) and support vector machine (SVM). | The hierarchical process model of CRISP-DM | LR is better than SVM to predict student performance. |
| 4          | 2018                | Santoso L W & Yulia | English | To compare multiple linear regression (MLR) and single linear regression (LR) in predicting student performance. | Multiple linear regression | MLR performs better than LR in predicting student performance. |
| 5          | 2011                | Sembiring S et al | English | To investigate if behavioral factors are predictors of student performance. | Decision Tree and Smooth SVM | Behavioral factors significantly influence student performance. |

In the following section, we present critical analysis for the published researches. The key attributes of the articles were identified along with our analysis on the articles.

3.1. Critical analysis
The first article was written by Mardiana and Faqih [6] and published in 2019 in Barekeng Jurnal Ilmu Matematika dan Terapan (Barekeng - Journal of Mathematics Science and Application). The article was written in Indonesian language. The title is “The best Structural Equation Modelling – Partial Least Square Model (SEM-PLS) to Evaluate A Learning Management System of Discrete Mathematics Course”. The research was motivated by the lack of studies that investigate the relationship between the quality of learning management system (LMS) with students’ final score [6]. The objective of the research was to get the best SEM-PLS model that can satisfactorily describe the relationship between LMS quality and students’ final score. A number of 109 third-semester-students participated in the study.

The independent variables in the study were: perceived of usefulness, perceived ease of use, and relevance of content learning. While the dependent variables were: interest of learning and quality of learning. Data for these variables were gathered through questionnaires. Using the variables, the main model was designed and evaluated. Then 16 alternative models were designed based on the main model.
All of the models were then evaluated to examine their validity, reliability, predictive accuracy ($R^2$) and predictive relevance ($Q^2$). The result showed that 2 models had the better $Q^2$ values. They are: the main model and an alternative model called C2. To choose between these 2 models, other criteria outside the $Q^2$ criteria were evaluated. And the result showed that C2 satisfied the criteria. Therefore, model C2 was selected as the best model to describe the relationship between LMS quality and students’ final score.

The study covered sufficient variables to design the model that describe student learning experience in an LMS environment such as students’ perceptions on usefulness, ease of use of the LMS, and self-regulated learning. These perceptions have been considered as important variables in other LMS studies [7,8]. The study firstly based their decision on $R^2$ and $Q^2$ criteria to select the best model because the predictive accuracy and relevance values are primary properties of a predictive model [9]. When presented with 2 final choices of models, the authors used 6 other criteria to narrow down the selection.

The second article was written by Alyani and Nurafni [10]. The article was published in 2019 in IOP conference proceedings and was written in English. The title is “Structural Equation Modelling (SEM) in Predicting Student Performance Factors in Mathematics Education Department at Muhammadiyah University of Prof. DR. Hamka”. The objective of the study was to investigate the influence of gender and educational background on students’ academic performance. 200 students participated in the study.

The procedure of the study was started with deciding the variables. The independent variables were gender, age and educational background (public, private or religious school); and the dependent variables were mid-test and final test. Students’ academic performance was based on students’ grade point average (GPA). Analysis using SmartPLS software showed that age and educational background did not significantly contribute to GPA. Therefore, these variables were eliminated from further steps. Next, the final model was analysed using SmartPLS. The final model was in the form of a diagram. It was concluded that gender, mid-test and final test are significant factors that influence students’ GPA.

The study included gender to predict student performance. This is reasonable because literatures, as listed by Joseph et al [11], seem to be in disagreement about the influence of gender on student performance: some studies claim that there is no significant influence, but others argue that boys perform better than girls, or otherwise. The study concluded that gender was indeed an influential factor in the prediction model. Unfortunately, it did not clarify whether male or female perform better.

The third article was written by Khasanah and Harwati [12]. The article was published in 2019 in the International Journal of Information and Education Technology. It was written in English. The title is “Educational Data Mining Techniques Approach to Predict Student’s Performance”. The objective of the study was to estimate student performance using Linear Regression (LR) and Support Vector Machine (SVM) models. 104 data set were collected from the information system of Universitas Islam Indonesia.

The method of the research followed the Cross Industry Standard Process for Data Mining (CRISP-DM) which is a hierarchical process model [12]. The first 3 steps dealt with describing the attributes and origins of data. The next step was the modelling step which included K-means algorithm for data clustering, and LR and SVM for classification models. The last step was evaluation and deployment where the estimation accuracy of LR and SVM was compared using root mean squared error (RMSE).

The result showed that clustered data were better for prediction than non-clustered data; and that LR outperformed SVM in predicting student’s academic performance [12]. Unfortunately, the study did not discuss whether their findings support or opposed to previous studies that have been conducted in this area. This information might be useful for other researchers in choosing which prediction model to use.

The study covered internal factors to predict a student’s performance at a university. The factors were: senior high school grade, attendance, GPA in semester 1 and final GPA [12]. The model might give better contribution for decision makings if external factors were also considered such as teacher motivation, access to books and other sources, the use of technology, and the interactions with the teacher and other students.

The fourth article was written by Santoso and Yulia [13]. The article was published in 2018 in the Journal of Telecommunication, Electronic and Computer Engineering. It was written in English. The title is “Predicting Student Performance Using Data Mining”. The objective of the study was to compare
the use of multiple linear regression (MLR) and single linear regression (LR) in predicting student performance. Data were gathered from the LMS of Petra Christian University. The data were categorized into student-centered features (GPA), activity-centered features (quiz or assignment and the difficulty level of the course), and Lentera features (student online interaction).

The study conducted analysis for MLR and LR models both with and without Lentera features. Result showed that MLR models performed better than LR models, and that the inclusion of Lentera features improved the accuracy of the prediction [13]. Unfortunately, this article did not discuss the comparison between their findings to other published works in the same area.

The fifth article was written by Sembiring et al [14]. The article was published in 2011 in the proceedings of the International Conference on Management and Artificial Intelligence. It was written in English. The title is “Prediction of Student Academic Performance by an Application of Data Mining Techniques”. The objective of the study was to investigate whether behavioral factors could be used to predict student performance. 300 students of University Malaysia Pahang participated in the study. Data were collected from questionnaires. The variables considered as behavioral factors were interest, study behaviour, engage time and family support. The results showed that behavioural factors were strong predictors of student performance [14].

The study conducted several steps starting with data preparation, data clustering using K-means algorithm, and data classification using Decision Tree and SSVM methods. In our opinion, the organization of the writing is rather hard to follow, especially for readers who are new to mathematical modelling. Also, there was no discussion on how the findings supported or opposed to other published works in the same research field.

3.2. Synthesis
From the critical analysis in the previous section, several points can be stated as follow. Firstly, there were only 5 articles that were discovered to be reviewed for the present study. This indicates that there is a big gap in the area of mathematical modelling for Indonesian educational research. Future research should make use of mathematical modelling because research findings around the world have proven that mathematical modelling is a powerful method to analyse student performance.

Secondly, the modelling methods that were used in the articles were Partial Least Square (PLS), Partial Least Square Structural Equation Modelling (PLS-SEM), Linear Regression (LR), Multiple Linear Regression (MLR) and Support Vector Machine (SVM). Future research on mathematical modelling in the context of Indonesian educational study can employ these models. But there are other methods to be considered as well. For example, multilayer perception network model and the radial basis function network model [15]. A researcher can use the method that is suitable for their research objectives.

Thirdly, mathematical modelling in the articles were basically used to predict student performance or to investigate the factors that influence student academic achievement. This is in line with previous studies conducted in other countries [6,7]. Future research can consider the use of mathematical modelling for other educational purposes, for example to develop a student feedback model or to predict the rate of job opportunities for fresh graduates.

Lastly, the variables that were included in the prediction of student performance in the articles were mostly internal factors such as student perception in article 1; gender and test scores in article 2; high school grade and attendance in article 3; GPA in articles 3 and 4; student interest, study behaviour and engage time in article 5. Future research can consider including other variables such as teacher motivation, teaching strategies, access to learning resources, the use of technology in learning activities, and the interaction pattern within the classroom.

4. Conclusion
In the introduction section, these following research questions were proposed: what were the objectives of using mathematical modelling in the published researches? what mathematical modelling methods were used in the published researches, and what are the prospects for future research in Indonesia
regarding mathematical modelling? Based on the review, these questions can be answered as follow. First, the objectives of the articles were to predict student performance or to investigate the factors that influence student performance. The mathematical modelling methods that were used in the articles were Partial Least Square (PLS), Partial Least Square Structural Equation Modelling (PLS-SEM), Linear Regression (LR), Multiple Linear Regression (MLR) and Support Vector Machine (SVM). Future prospects in Indonesia are the implementation of various mathematical modelling methods for various educational purposes since there has been very few publications in this field; and the inclusion of more extensive factors or variables in modelling student performance to incorporate comprehensive internal and external factors that may influence student academic achievement.

References
[1] Haines C and Crouch R 2007 Mathematical modelling and applications: Ability and competence frameworks In W Blum, P L Galbraith, H Henn & M Niss (Eds) Modelling and Applications in Mathematics Education: The 14th ICMI Study pp 417-424. New York: Springer
[2] Lawson D and Marion G 2008 An Introduction to Mathematical Modelling (Bioinformatics and Statistics Scotland)
[3] Musso M, Kyndt E, Cascallar E and Dochy F 2012 Predicting mathematical performance: The effect of cognitive processes and self-regulation factors Education Research International 2012 1-13
[4] Li M, Fang Q, Cai Z and Wang X 2011 A study of influential factors in mathematical modeling of academic achievement of high school students Journal of Mathematics Education 4 31-44
[5] Torraco R J 2005 Writing integrative literature reviews: Guidelines and examples Human Resource Development Review 4 356-367
[6] Mardiana N and Faqih A 2019 Model SEM-PLS terbaik untuk evaluasi pembelajaran matematika diskrit dengan LMS [The best SEM-PLS model for discrete mathematics lesson evaluation with LMS] Barekeng 13 157-170
[7] Ghazal S, Aldowah H, Umar I and Bervell B 2018 Acceptance and satisfaction of learning management system enabled blended learning based on a modified Delon-McLean information system success model International Journal of Information Technology Project Management 9 52-71
[8] Eom S B 2012 Effects of LMS, self-efficacy, and self-regulated learning on LMS effectiveness in business education Journal of International Education in Business 5 129-144
[9] Hair J F, Hult G T M, Ringle C M and Sarstedt M 2017 A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM) 2nd ed (California: Sage Publication)
[10] Alyani F and Nurafni 2019 Structural equation modelling (SEM) in predicting student performance factors in mathematics education department at Muhammadiyah University of Prof. DR. Hamka J.Phys: Conf. Ser. 1315 012040
[11] Joseph A, John O, Eric I, Yusuf S and Olubunmi A 2015 Effect of gender on students’ academic performance in computer studies in secondary schools in New Bussa, Borgu local government of Niger state Journal of Education and Practice 6 1-7
[12] Khasanah A U and Harwati 2019 Educational data mining techniques approach to predict student’s performance International Journal of Information and Education Technology 9 115-118
[13] Santos L W and Yulia 2018 Predicting student performance using data mining Journal of Telecommunication, Electronic and Computer Engineering 10 1-4
[14] Sembiring S, Zarlis M, Hartama D, Ramliana S and Wani E 2011 Prediction of student academic performance by an application of data mining techniques Proc. Int. Conf. on Management and Artificial Intelligence (Bali, 1-3 April 2011) pp 110-114
[15] Huang S and Fang N 2013 Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models Computers & Education 61 133-145