Mapping of Students’ Academic Performance in Online Learning Environment during Pandemic Using Multiple Correspondence Analysis

Ainil Fauzani Rosmadi, Shazlyn Milleana Shaharudin*, Murugan Rajoo, Rawdah Adawiyah Tarmizi, and Mohd Saiful Samsudin

Abstract—The virtual and digital learning process seems to hugely impact academic achievement due to the COVID-19 outbreak, globally. Thus, improving student performance is one of the important focuses of educational management. Mapping students’ actual conditions is a mandatory requirement before designing the performance improvement program. Therefore, this study proposed a statistical investigation to map out students’ performance and the problems they encountered during online learning using Multiple Correspondence Analysis (MCA), revealing the hidden pattern and classifying students based on their demographics (programs, CGPA, and origin) and learning environments. The data samples consist of 234 undergraduate students in the Department of Mathematics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris (UPSIP). The study findings formed from two different profiles where each profile has its respective categories. The study found that the students of the Bachelor of Education (Mathematics) with Honour, categorized as smart students, preferred to study face-to-face because of poor internet connection from using mobile data. On the other hand, the students of the Bachelor of Science (Mathematics) with Education, who were categorized as average students, had no difficulty continuing either synchronous or asynchronous online learning in the future because of stable internet access using their home Wi-Fi connection. Moreover, the preference made was also due to family interruption issues.

Index Terms—Mapping, academic performance, learning environment, multiple correspondence analysis.

I. INTRODUCTION

Coronavirus disease 2019 (COVID-19) was firstly identified in Wuhan City, Hubei Province, China in December 2019 where COVID-19 is a potentially fatal viral disease caused by the deadly Corona Virus, that has been spread around the world. Therefore, the World Health Organization (WHO) has classified it as a global issue and announced it as a pandemic on 12th March 2020. Malaysia is also not exempted from the pandemic. Thus, several governmental measures had been taken to counteract the risk of disease. These measures included travel restrictions, mandatory quarantines for travellers, social distancing, school and university closure, bans on public gatherings, business closures, self-isolation, curfews, and lockdown. On 16th March 2020, Tan Sri Muhyiddin Yassin, the Prime Minister of Malaysia declared the Movement Control Order (MCO), which was announced to be implemented from 18th March 2020 to 31st March 2020 [1]. The pandemic has dramatically reshaped the way global education is delivered. Millions of learners were affected by educational institution closures due to the pandemic resulting in the largest online movement in the history of education. With this sudden shift of being away from classrooms in many parts of the globe, universities had also rapidly shifted to virtual and digital strategies.

Universiti Pendidikan Sultan Idris (UPSIP) correspondingly closed the vicinity and announced restrictions for students and lecturers from attending face-to-face classes. Due to the suspension of classroom teaching, the learning process switched to online learning, which continued for months. This form of learning provides an alternative way to minimize the contact between students or between the students and lecturers [2]. Furthermore, the quality of education and excellent infrastructures such as computers and IT modern equipment reception are now in massive demand and universities are changing their teaching models with the use of intellectual capital [3]. However, many students have no access to online teaching due to a lack of means or instruments because of the economic and digital divide. Hence, these virtual and digital strategies as a whole learning process seem to give a huge impact on university students, especially on their academic achievement [4].

A significant problem of online learning is that its entirety is dependent on technological devices and the internet. Hence, instructors and students with poor internet connections are liable to be denied access to online learning. A. Nguyen [5] stated that students faced many obstacles in a home learning environment, such as lack of mastery of technology, high internet costs, and limited interaction or socialisation between and among students. This would often happen to students living in rural areas with poor internet connections, which unquestionably interrupted the whole learning process and changed the mood. Other than that, poor internet connections caused more problems for students, such as affecting their attendance in class, which led to misinterpreting what they learned. In addition, they also had difficulty receiving any learning materials shared by their lecturers as well as submitting their work, leading to declining academic achievement. M. Suryaman et al. [6] investigated that lockdown made significant disruptions in students’ learning experience. The students also reported some challenges that they faced during online classes. Besides, students’ learning

*Correspondence: shazlyn@fsmt.upsi.edu.my

Manuscript received April 16, 2022; revised July 12, 2022; accepted July 22, 2022.

Ainil Fauzani Rosmadi, Shazlyn Milleana Shaharudin, Murugan Rajoo, and Rawdah Adawiyah Tarmizi are with Department of Mathematics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, Malaysia.

Mohd Saiful Samsudin is with Environmental Technology, School of Industrial Technology, Universiti Sains Malaysia, Malaysia.
environment is also vital for the students to be more focused and becomes productive in class. Some studies indicated that the home environment can affect students’ academic achievement. For example, parents’ constant disagreement affected their children emotionally and this could lead to poor academic performance [7]. Some students came from broken homes, which made them prefer living away from their families, and the university became the best place for them to stay. Due to the closure of universities, most students were sent back to their hometowns. Moreover, problems in the home environment can also be a challenge in this online learning, such as unexpected appearances or interruptions of family members or pets that may cause disruption and diversion of online learning participants’ attention during the online learning process [8]. Therefore, the current study was conducted to identify the problems that the students encountered during online learning by mapping out the problems and academic performance of Universiti Pendidikan Sultan Idris (UPSI) Mathematics undergraduates using Multiple Correspondence Analysis (MCA) to reveal the hidden pattern and classify students based on their demographics (programme, CGPA, and origin) and learning environment during online learning.

Moreover, this study intends to provide an overview, especially in the educational field, for instance, the Institute of Higher Education of the aspects that can affect students’ academic performance such as student background and learning environment. Hopefully, the current study enables to provide guidance for educational management and policymakers as both groups seek to understand better and address the inherent challenges of enhancing students’ academic performance if online learning is continued. Last but not least, the result from this study hopefully can provide many beneficial things to certain parties who are involved to give better solutions and suggestions to improve the problems that students encountered if online learning is continued in order to improve academic performance.

Investigation of factors related to the academic performance of university students become a topic of growing interest in a higher educational circle. Many recent studies were carried out to explore factors that affect university students’ academic performance. Moreover, students’ academic performance is affected by several factors, which include students’ learning skills, parental background, peer influence, teachers’ quality, and learning infrastructure [9]. Issues like gender differences, teachers’ education, and teaching style, class environment, socioeconomic factors, and family education background can affect students’ academic achievement [10]. The findings of this study vary from region to region and the results differ in cities and rural areas. Education is an exceptionally vital issue as it influences the development of a country. The main objective of higher education institutions is to present quality education to their students. This is because education has the power to induce change and progress in society. A researcher carried out a study with university students in Baghdad to explore the factors that affect students’ academic performance and they found that environmental, economic, social, and psychological had a strong influence on students’ performance [11] whereas another researcher conducted a study to find the factors affecting undergraduate students’ performance in Bangladesh. In their study, they focused on exploring the factors associated with the performance of students through their primary data [12]. The research concluded that the family background, educational environment, and financial conditions of the learners were the crucial factors affecting academic performance, along with motivation and relationships between teachers and students that were directly related to achieving their academic goals.

Improving academic performance necessitates knowledge of the students’ condition. Mapping of students’ actual conditions is the requirement that must be made before designing the performance improvement program [13]. This is because students differ in their levels of motivation, understanding of teaching and learning, and environmental conditions. Several studies were conducted to map the students’ performance using a variety of variables and methods such as the Naive Bayes classification algorithm [14], the decision tree algorithm, and the K-means cluster algorithm [15]. The findings show that data mining techniques were not only useful for mapping and classifying students based on demographic variables, understanding of the learning process, and level of activity, to name a few. They could also be used to predict student achievement levels such as GPA, test scores, and the duration of the study depending on the suitable data mining techniques. However, in this study, the researchers only use Multiple Correspondence Analysis (MCA), a procedure that often appears to be the counterpart of Principal Component Analysis (PCA) for categorical data. MCA has the ability to apply intuition to a large dataset using detailed visualisation functionality that is built into the method [16]. In order to map out the performance and problems encountered by students during online learning, and to classify them based on their demographics and learning environment, the conceptual framework is shown in Fig. 1.

II. MATERIALS AND METHODS

The current study is focused on the application of Multiple Correspondence Analysis (MCA) to analyse a dataset associated with students’ academic performance and their learning environment during online learning even though there is no substantial literature on statistical methods in students’ academic performance regarding this MCA method. However, the researchers acknowledged that there are a considerable number of studies on students’ performance [17]. For instance, some researchers used the MCA method to
investigate students’ engagement indicators for academic performance. MCA achieves coherent analysis by grouping indicators according to their similarity. Accurate results cannot be obtained simply by employing a correlation-based method or a traditional regression analysis scheme. A common finding in the literature is the difference in tools used by researchers to assess student performance. Nonetheless, explaining students’ performance, which is not the focus of this study, can be intricate because it is a complex construct influenced by multilevel factors. Furthermore, the literature showed unclear students’ performance factors that have a direct impact on academic performance. Therefore, this study only focuses on mapping students’ academic performance in an online learning environment by identifying the students’ problems encountered during online learning and how the problems were related to students' performances using the MCA method to reveal the hidden pattern. By mapping students’ performance during online learning, a researcher can observe trends of students’ performance before the commencement of COVID-19 outbreaks and during online learning. This is due to the fact that numerous aspects can exist when online learning is adopted, causing some good and bad effects on students’ performance [18]. Thus, by mapping out the students’ performance, problems encountered, and preferences during online learning implemented, related parties can take initiatives to improve academic performance even when online learning is conducted.

A. Research Design

The design of this study is in the form of a qualitative approach in which the researchers used categorical data. Multiple Correspondence Analysis (MCA) is one of the methods that can be used for analysing categorical variables and obtaining a general understanding of how categorical variables are related. MCA method can be used for more than two variables in which multiple correspondences can be applied to a table with more than two dimensions. However, with more than six to seven variables, the resulting maps become challenging to use because it plots every level of each categorical variable. Often, this means that redundant information is plotted. In this study, the MCA method is used to see the relation between students’ problems encountered in the online learning environment and their academic performance. This research is carried out at Universiti Pendidikan Sultan Idris (UPSI), located in Tanjong Malim, Perak.

B. Data Collection

The data was collected through the research instrument of this study - a questionnaire. In the current study, the data collected were based on the questionnaire which is students’ demographic (programme, CGPA, and origin), students’ problems encountered, internet accessibility, and students’ preferences during online learning as the categorical variables. A total of 600 Mathematics students from the Faculty of Science and Mathematics were chosen for this study and the samples were randomly selected, which are 234 students from Semester 3 to 7 consisting of 43% male and 57% female, as shown in Fig. 2. The respondents who participate in the study considerable undergo online learning throughout the pandemic COVID-19. The researcher focused on Mathematics undergraduate students as samples with two types of Mathematics Programmes in UPSI, which are Bachelor of Education (Mathematics) with Honour (the programme code is AT14) and Bachelor of Science (Mathematics) with Education (the programme code is AT48). Both programmes have significant similarities; the difference is that the AT14 undergoes practical teaching at the end of their semester while the AT48 undergoes industrial training. 59.4% of them were AT14 students and 40.6% were AT48 students are shown in Fig. 3, with total respondents of 234 students.

![Students' Gender](image1.png)

**Fig. 2. Semester 3 to 7 mathematics undergraduates’ gender.**

![Programmes](image2.png)

**Fig. 3. Mathematics undergraduates based on programmes.**

C. Data Analysis

Multiple correspondence analysis (MCA) is a descriptive method that is an extension of correspondence analysis (CA) and allows you to investigate the pattern of relationships between several categorical dependent variables. Furthermore, MCA can be considered a generalisation of principal component analysis for categorical variables, revealing patterning in complex data sets. According to [19], MCA is useful to map both variables and individuals, allowing the construction of complex visual maps whose structuring can be interpreted. It is performed on an $N \times K$ indicator matrix in which $N$ is the number of data samples, and $K$ is the number of features describing the samples. The element in the cell $(n, k)$ of the indicator matrix consists of individual information $n$ and category $k$ [20]. Related categories in MCA are found close together in Euclidean space, leading to clouds of data points that have comparable distributions [21]. Remarkably, MCA’s output generates
two-point clouds that are typically represented by a two-dimensional graph. A two-dimensional plot gives more information about correlations between variables than higher-dimensional ones. In general, MCA can be defined as the application of Principal Component Analysis (PCA) to the indicator matrix [19]. The cloud of individuals is constructed on distances between individual information for an indicator, for which diverse categories of indicators have been selected [20]. Meanwhile, the squared distance between individuals related to each category is obtained in the case of each indicator [21]. An initial descriptive statistical analysis was performed to report the modalities of each indicator in the same direction. In this study, MCA was performed using XLStat 2019 Version as the tool to map out performance and problems encountered by the students during online learning and classify them based on the Symmetry Variable Plot obtained from the statistical data analysis.

III. RESULTS AND DISCUSSIONS

The MCA method was performed using XLStat 2019 Version as the statistical tool. XLStat can perform various data analyses, presentation of functions, statistical analyses, and graphical data depiction.

A. Mapping of Students’ Performance, Problems, Accessibility and Preferences during Online Learning

According to [23], the magnitude information related to each dimension is termed eigenvalue with 0 and 1, which this study indicates the total variance among the problems encountered and demographic of students. Every point on the plot contributes to all dimensions, and the scale of the plot is heavily influenced by the volume of contributions from each dimension. This study discovered that the first and second dimensions had a higher eigenvalue than other dimensions. The first and second dimensions had eigenvalues of 0.289 and 0.217, respectively presented in Table I. Together, these two dimensions accounted for approximately 32.20% of the data variability. Furthermore, the low eigenvalues for the data set demonstrate the heterogeneity of the problems encountered. This heterogeneity may be due to the questionnaire measures of randomness.

![Symmetry variable plot.](image)

Based on the scree plot presented in Fig. 5, the model estimation had resulted in the formation of 10 principal dimensions. The MCA method estimated the number of dimensions to be formed, and each dimension explains a certain amount of variance within the questionnaire data set. For instance, dimension 1 explains 18.38% of the data sets’ total variance while dimension 2 explains 13.82% of the data sets’ total variance and the total inertia for the principal dimensions only took 1.571. Therefore, based on the Symmetry Variable Plot in Fig. 4, axes F1 (dimension 1) and F2 (dimension 2) carry 32.20% of the survey information and are explained as the output.

![Scree plot.](image)

B. Classification of Demographics of Students and Their Learning Environment

Various classification techniques can be used to obtain the desired study results. However, in this study, the researcher used Multiple Correspondence Analysis (MCA) to classify students based on their demographics and learning environment during online learning. The researcher identified the issues that students encountered during online learning by mapping the issues, as shown in Fig. 4. Then, the researcher classified the students based on their demographics and learning environment to provide appropriate learning strategies for students in the future. Hence, as shown in Fig. 4, the MCA output allows a clear distinction between two...
profiles:

- **Profile 1:**
  Which is in a blue colour box, are categorised as smart students and most of them are in the first class of CGPA. They are students from the Bachelor of Education in Mathematics with Honour (AT14) programme who mostly reside in suburban or rural areas. They have a large family of 7 or more members. The problems encountered by the students during online learning are due to the unstable internet connection. This is most likely because of the use of mobile internet data. Therefore, they preferred to study face-to-face in the future.

- **Profile 2:**
  Which is in a green colour box, are categorised as the average students and most of them are in the second class of CGPA. They are students from the Bachelor of Science (Mathematics) with Education (AT48) programme who mostly reside in urban areas. They have a small or average family size of 1 to 6. The problems encountered by the students during online learning are due to family interruption, sound pollution, and part-time jobs. Their internet accessibility is their home Wi-Fi connection. Hence, they preferred to study through blended learning (offline online learning) or online learning, which are either synchronous or asynchronous.

C. **Discussions**

This study was designed to identify problems that students encountered during online learning associated with their learning environment during a pandemic and to classify whether they are subgroups with similar pattern profiles. Additionally, this study highlights the students’ problems during online learning that mostly contribute to variance in the MCA analysis and their main correspondences. The classification obtained could be considered in the future to provide appropriate learning strategies for students. Thus, in this phase, the researcher discusses the classification obtained where two profiles were explained in the previous section.

The attributes used as inputs in this research are the data of Mathematics undergraduates in Universiti Pendidikan Sultan Idris (UPSRI) in semesters 3 until 7 consisting of programme, CGPA, origin, family background, internet accessibility, disturbance during online learning, and study preferences. The amount of data used is 234 Mathematic students who formed two different profiles where each profile has its respective categories. As shown in Fig. 4, there are two boxes. Profile 1, represented by the blue colour, is students from the Bachelor of Education (Mathematics) with Honour (the programme code of AT14), categorised as smart students who are in the first class of CGPA, whereas Profile 2, represented by the green colour, is students from the Bachelor of Science (Mathematics) with Education (the programme code of AT48), also categorised as average students who are in the second class of CGPA. The study found that the AT14 students preferred to study face-to-face because of poor internet connection from using mobile data while the AT48 students preferred to study through online learning either in synchronous or asynchronous method because of the fast internet connection using their home Wi-Fi. Additionally, students’ preferences were also caused by their family interruption. The preference was made based on the mapping shown in Fig. 4 where family interruption is the closest point to the other points of sound pollution and part-time jobs. The categorisation between the blue and green colour boxes was made due to the results of the MCA algorithm in a two-dimensional plot (x-axis and y-axis representing dimensions 1 and 2 respectively) that shows the correlation between the variables. As we can see in Fig. 4, the points of programme-AT14 are very close to other points, which are internet accessibility-mobile data, CGPA-1st Class, living location-suburban and rural area, disturbance during an online learning-slow internet connection, family size-7 and above, and last but not least, study preference-face-to-face (offline). These were categorized in the blue colour box, which defined the correlation within the variables in that box. The same steps were implemented into the green colour box. This analysis assists the researcher in understanding the problems encountered by students while undergoing online learning by distinguishing the two groups that occur as a result of the analysis. Consequently, the researcher can create reform measures in the university education system in this endemic era where we still need to be vigilant with epidemics that have occurred and must adhere to the standard operating procedures established by the government.

One of the most difficult challenges in addressing the problem of online learning is the availability of technological gadgets and internet connectivity. There were pieces of literature that supported the argument of this study and saw the needs and challenges of internet connections among students [24]-[27]. This is the main problem that AT14 students faced when using mobile internet data for internet access in online learning. This may also be related to their current residence, as most students live in suburban and rural areas. According to [28], almost 35% of the population still does not have access to steady internet service, especially in rural areas, either because of poor coverage or because the people just cannot afford it. He also pointed out that access to the internet has never been more crucial than now given the restriction on normal life brought about by the COVID-19 pandemic. Therefore, AT14 students would rather study face-to-face for future learning in order to address problems and improve their academic achievement.

While on the contrary for AT48 students, the main problems they faced during online learning were because of family interruptions. According to [11], unexpected interruption of family members or pets may cause disruption or diversion of online learning participants’ attention during online learning, such as house chores given by parents, e.g., babysitting for younger siblings. Tasks such as babysitting or being called to work affected students’ attendance and participation in scheduled online lectures [29]. However, AT48 students had no difficulty continuing online learning either in synchronous or asynchronous for future learning as they used their home Wi-Fi for internet access. Moreover, to overcome the problem, they need to manage their time wisely or make a timetable for house chores. Besides time management, students would also need educators’ lenience, such as giving an appropriate time to complete assignments. As noted by [30], [31], some universities offer asynchronous classes where instructors prepare assignments or record lectures, and students can complete them at their own pace.
IV. CONCLUSION

In this study, the researcher used Multiple Correspondence Analysis to analyse students’ academic performance associated with the learning environment during online learning. The main objective of this study is to identify students’ problems encountered during online learning by mapping students’ performance, problems, accessibility, and preferences during online learning. Apart from that, classified students based on their demographic such as programme, CGPA and origin, and their learning environment during online learning. From the results obtained, the MCA output clearly distinguished two different profiles. Profile 1, which is AT14 students categorised as smart students, had slow internet connection problems due to their current location as most of them live in suburban and rural areas. For Profile 2 where the AT48 students were their current location as most of them live in suburban and smart students, had slow internet connection problems due to the use of mobile internet data. This could also be related to their current location as most of them live in suburban and rural areas. For Profile 2 where the AT48 students were categorised as average students, the problems encountered were due to family interruption, but they had no difficulty continuing either synchronous or asynchronous online learning for future learning as they used their home Wi-Fi with a stable internet connection. In the future, this research can be extended by processing the same variables and adding more detailed information to predict students’ academic performance during online learning by using classification prediction techniques. Hopefully, this research can provide beneficial input, especially for educational management to improve students’ academic performance as well as helping students who are in need.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Conceptualization, S.M.S.; methodology, A.F.R.; software, M.R. and A.F.R.; validation, R.A.T. and M.S.S.; writing—original draft preparation, A.F.R. and S.M.S.; writing—review and editing, A.R.F. and M.R.; project administration, S.M.S. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGEMENT

The authors would like to thank the Ministry of Higher Education Malaysia (MOHE) for supporting this research under Fundamental Research Grant Scheme Vote No. 2019-0132-103-02 (FRGS/1/2019/STG06/UPSI/02/4).

REFERENCES


Ainil Fauzani Rosmadi was born in Selangor, Malaysia. She is a bachelor degree student in science (mathematics) at Universiti Pendidikan Sultan Idris (UPSI). In her studies, she focused on mapping and classifying students' performance and problems' encountered during online learning using multiple correspondence analysis.

Shazlyn Mylleana Shaharudin was born in Johor Bahru, Malaysia, in 1988. She is a senior lecturer at the Department of Mathematics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris (UPSI). She graduated with a bachelor science degree in industrial mathematics from Universiti Teknologi Malaysia, in 2010. Upon graduation, she began her career as an Executive in banking institution. In the following year, she received an offer to continue her study as a fasttrack PhD student at the same university. During her PhD journey, she developed an interest in multivariate analysis, specifically in finding patterns which deals with big data. Her research focuses on the area of dimension reduction methods specifically in climate informatics which involves analysis on huge climate-related datasets based on techniques in Data Mining. She had published her research in Scopus indexed journal and presented her work in various local and international conferences. She completed her PhD thesis at the end of 2016 and was conferred a doctorate degree in 2017.

Murugan Rajoo is currently working as the senior lecturer at Sultan Idris Education University in Malaysia. He obtained his Ph.D. in mathematics education from University Utara Malaysia (UUM). He was also appointed as the head of the Olympiad Mathematics Competition (2011 – 2013) and Kangaroo Mathematics Competition (2017 – 2018). Additionally, he was invited as a speaker to the iSTEM International Conference organized by Universiti Sains Malaysia (USM, 2019). Besides, he also coordinated a workshop on computational thinking. Dr. Murugan was involved in STEM module and instrument development for the teachers professionalism division at the ministry level. His works have been published widely, including ISI and Scopus journal. He also contributes to expertise by being a reviewer and editor for International Journals. Dr. Murugan has won many awards and recognitions. His areas of expertise are mathematics education, assessment in mathematics, STEM education, mobile applications, computational thinking, and instrument development. Currently Dr. Murugan actively engages in research on promoting global citizenship education.

Rawdah Adawiyah Tarmizi was born in Ipoh, Malaysia, in 1986. She is currently a senior lecturer at the Department of Mathematics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris (UPSI). She graduated with a bachelor's science degree in mathematics from Universiti Teknologi Mara, in 2008. Then, she completed her M.Sc. in mathematics at Universiti Kebangsaan Malaysia in 2011. She holds a PhD in mathematics from Universiti Sains Malaysia Pulau Pnang specialization in Algebra and Group Theory. She has experience teaching in the field of mathematics after joining UPSI as a tutor in 2009. Besides teaching, she is actively doing research in her field and mathematics education.

Mohd Saiful was born in Pahang, Malaysia, in 1988. He is a senior lecturer at the environmental technology, School of Industrial Technology, Universiti Sains Malaysia. He graduated with a bachelor science degree in environmental science & his master degree in environmental analysis & modelling from Universiti Putra Malaysia, in 2010 & 2014. In 2019, he completed his doctorate at Universiti Sultan Zainal Abidin majoring in applied science (environmental toxicology/chemometrics).