Model of Education Technology for Language Pedagogy in Higher Education

Risa Simanjuntak1,2*

1English Department, Faculty of Humanities, Bina Nusantara University, Jakarta, Indonesia 11480
2Digital Language and Behavior Research Interest Group, Bina Nusantara University, Jakarta, Indonesia 11480

Abstract. Education technology enables advances in every aspect of education. This paper explored a model for language pedagogy through Educational Data Mining (EDM). EDM has offered important contributions in the last decade. With EDM, many predictions could be made in terms of learning paths, patterns for success and failure, and students’ preferences. Such predictions would be much needed for decision-making, business, and academic-wise. However, not enough EDM has been done regarding language learning. This present study provides a potential model for EDM in language pedagogy. A substantial review of the literature was complemented with samples of data from students’ language learning performance as illustrations for the model. Corpus for this study was students’ writing from various universities. Results showed the need to integrate language input, process, and output into EDM and create a base model of learning. Predictions for learning challenges, problems, and failures would be beneficial to improve teaching and learning. In conclusion, EDM was inevitably needed in the rise of online learning. Practical implications for language platforms and digital language learning were also discussed.

1 Introduction

Educational Data Mining (EDM) is an emerging field of study, which takes data mining to further predictions for educational purposes [1-3]. As a multidisciplinary field of research EDM is used to analyze educational data using data mining techniques [1]. EDM classifies, extract, and display information from big data to answer educational questions [2]. EDM employs data mining techniques and tools and apply them for gaining strategic information. Such information needs to have valuable impacts and turn-out. Especially in business, strategic information could influence business decisions and ventures [1-3].

Language education post COVID-19 pandemic has taken a sharp turn. This has been influenced by learning routes via online platforms [4, 5]. Language pedagogy has benefitted a lot from the provision of online, hybrid, and blended learning pathways. With the junctures, activities of selecting, navigating throughout learning, monitoring, and assessing learning have much been done online. Online activities become the core actions in language exchanges. Unlike classroom interactions which largely are not recorded and stored, language activities online often were recorded, stored, and shared. The amount of data online was massive and largely individual rather than communal. Also, an individual learner could have many samples of their language performance from the language used online. Samples of performance build patterns of learning evidence of an individual learner. Certainly, in higher education level, online learning and hybrid mode provided by the universities during pandemic time provides lush amounts of data on learning.

EDM in language pedagogy would be beneficial to prepare learners for real communication opportunities and challenges [1-3]. With EDM not only evidence of learning could be read but more specifically in language pedagogy, EDM would gather data of language performance online. In online communications, corpus is built from online texts. The texts can be compiled into reference corpus, or model corpus, which is made up of language use samples from professional communities. With the provision of reference corpus, language pedagogy would benefit in using the patterns of language used by professional and industrial contexts. Learners could use these patterns and language styles used in the reference corpus as the models for language learning. Comparing the reference corpus with the learner corpus provide important information for both the development in learning and further needs for learning [6, 7]. Using the data gathered online on language used by learners EDM could have powerful prediction of language success.

In language pedagogy evidence of learning could be used two-folds. First, evidence of learning reveals the learning performance and achievements of the learning in the past [6]. Second, evidence could also reveal further needs in learning [7]. The second would be useful to pave the learning path for an individual learner. Learning path in online learning is very important since a learner would need to interact with the prepared virtual environment. Evidence in the virtual environment would be beneficial for learners to apply in real

* Corresponding author: risa.simanjuntak@binus.ac.id

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).
communicative interactions. In the lights of linguistic analysis and pedagogical perspectives, it is in dire needs that EDM develops a more comprehensive model.

This present study aims at: (1) identifying essential dimensions in EDM for language pedagogy in higher education level, and (2) proposes an integrated model of EDM for language pedagogy in higher education.

2 EDM and language pedagogy

Technology in education enables the application of big data for acquiring large amount of information on language learning. Such provision of relevant and accurate information would be vital for decision making for language pedagogy and language policy in education.

2.1 EDM developments

EDM is largely defined as “...a group of dissimilar area that incorporates student result prediction and classification by using some techniques” (p.11) [1]. As a multidisciplinary field, EDM requires a comprehensive view on the data mining to present most accurate and effective predictions [1-4]. In EDM not only data is important but also the contexts and producer of data would equally need to be acknowledged. Previous studies in EDM have shown developments in models and applications of EDM [8-15]. Models of EDM, as presented in the literatures, acknowledge the needs of education and business applications [16-18]. These models also acknowledge the importance of metadata in learning patterns and predictions. However, previous studies did not specifically present a model for language skills, which require evidence outside the classroom setting. Models of EDM commonly use techniques, such as classification, clustering, and prediction. New developments in EDM also incorporates association rule mining [1, 3, 4]. Classification provides data with information needed by educators and learners. Typically, EDM with the purpose of predicting students’ success or failure in a course would require learning performance as part of the classification. Clustering enables the data to be grouped with similar attributes. Data with the same attributes would be clustered into the same group. The aim for clustering is to provide the necessary facilitation for the intended group. In the case of lower achievers, score would be an attribute [16-18]. With clustering, facilitation would be provided for the needed students and learning improvement could be expected. Prediction is a technique most desired to provide EDM impactful outcomes. With prediction unknown attribute could be identified with the help of known attributes. Prediction provides information from data being gathered in data mining.

Another technique is association rule mining [1, 3, 4]. This technique enables hidden attributes to be revealed in the same datasets by way of association. This technique is increasingly preferred in EDM as it relates different attributes and associate their relations. Different attributes were also known to be initially hidden. By way of association such relations could be revealed [1, 3, 4]. In the case of a student’s success in a university course, attributes to be collected could be more than just scores. Time allocated to learning, participation in forum, or perception on learning could be attributes. Previous research also found out through association rule mining that poverty and low motivation contributed to students failing their studies [19, 20].

Previous studies in EDM provided insights to education. Studies were predicting the schools’ drop-out and such was important to prevent one from happening [14, 16, 17, 21-23]. Models for EDM was also important to facilitate learning [24, 25]. The studies predicted students’ success in learning [26, 27]. Studies also looked at how virtual learning would help students learn [24-27]. EDM also was useful to generate students’ papers [26] and created learning environment even for distance learners [26, 28]. However, there had not many studies in language pedagogy’s EDM. Previous studies on language learning have not clearly exposed models incorporating dynamic events in communication. Language was seen as an object learnt rather than a skill mastered in that different genres in communication are not integrated as attributes in the EDM models. There is a clear gap between EDM and language pedagogy.

Studies in EDM at higher level education identified several attributes and clustered the data to present predictions of students’ success or failure. Most of the attributes, however, were labelled by the researchers. In language pedagogy there were factors affecting language performance, which are not prescribed by instructors or curriculum designers. Prescriptive rules are regulating teaching and learning in language courses in universities. Prescriptive rules, however, had been challenged by studies to be misleading and irrelevant to the dynamics of language use [14, 24].

2.2 Corpus-based analysis in language pedagogy

Corpus-based analysis relied on real examples of language use [14, 24]. Corpus, or collection of texts, can reveal the real use of language and the patterns learners have when performing in language. A corpus could also reveal the strategies of the language learners in communicating. Further analysis of the corpus could also reveal the identity of the learners. Corpus-based analysis has been used in complement for expert judgement of the language instructors. The use of corpus in learning assessment is useful to avoid bias and cross-cultural interpretation from the instructors. A study, for example, revealed that students may use cyclical style of when writing their academic papers [24]. Such nonlinear style was proven to be a portrayal of the first language narrative culture rather than low performance in language skill [7, 9, 14].

Corpus-based analysis is also used to reveal patterns and conventions of certain language communities. Previous studies revealed the use of inclusive we in writing research article as a known style in Computer Science [7, 24]. It was also revealed in another study that
a common syntactical structure of S-V-O with copula be
to present facts to be preferred in research publications
[24]. The finding of the study was similar with corpus
from other exact sciences, where simple sentence
structure was used rather than complex or compound
sentences [19, 24]. Corpus-based analysis confirmed the
notion that there is not only one model of language
which could be used by learners [23, 24]. And since
there was not one model-fits-all use of communication
purposes, a learner should not be assessed based on one
single model. Variations of communication model could
be applied by a learner and needs to be appraised
equally.

Variations in language used are necessary in
learning. When exposed to the varieties learners would
be able to experience the real communication
experiences [19]. Such variations or genre in
communication were different communicative events
[19]. Genre in language variety has patterns,
conventions, and style. Samples of genres range from
abstract of a research paper, email, cover letter, up to
movie and book reviews. These are all different
communicative events, which learners may encounter,
in real and daily communication needs. It is therefore
necessary to address the needs of learners to exposure to
genre.

Previous studies addressed the importance of genre
in the language pedagogy [19, 23, 24]. The knowledge
of genre was also important for bridging the language in
professional challenges for university’s graduate.
Exposure to genre was also proven to be effective to
enhance learning and learners’ confidence. The later
was essential for acquiring a new language [19, 23, 24].

3 Methods

Methods for this model of exploration was corpus-based
analysis and data mining, following a model previously
provided in Figure 1.

![Diagram](https://example.com/diagram.png)

**Fig. 1.** A three-axial model of language pedagogy in higher
education [25].

As can be seen from the figure, language pedagogy
in higher education is a dynamic environment. Ideally,
information on the ecosystem be integrated in the
learning [25]. This model for language pedagogy
encompasses industry, which becomes the outcome of

4 Model for language pedagogy in
higher education

The proposed framework for EDM for language
pedagogy in higher education can be seen in Figure 2.

The model presented EDM in an ongoing cycle of
identification, prediction, and re-identification. The
intention of the EDM is not to measure learner’s
distance from the point of completion. Such typical
EDM would see the end of the prediction to be to finish
the language course. On the contrary, language is a life
skill, and the nature of the skill is dynamic. Therefore, a
more organic EDM model is presented.

With the model data gathered, classified, and
associated were aimed at predicting the progress and
directions of the language learning. In the model, there
is a continuous process of identification, performance,
and redirecting the learning.

The model uses EDM to assists learning. Typical
EDMs provided information on whether learning would
fail and possible factors causing the low performance.
Information generated from the EDM models would
then try to prevent the occurring or reoccurring of
the factors. However, the present study offered a new model
of EDM, in which prediction would not be the final aim
of the data mining. Rather, when prediction was made,
the redirections for learning would be provided.

The model of EDM for language learning in higher
education was built on the notion that learning is
dynamic and ongoing. Language learning in higher
education is also aimed at training learners to tap in to
professional and industrial-ready graduates. Therefore,
the system of learning and curriculum design is closely
related to the language used in professional and
industrial contexts. The aim of the learning was not the
acquisition of the language but the continuous use of the
language and experiences in using the language. This
was based on the philosophical foundation of
knowledge being co-constructed and learning as
collaborative works.

The model uses three dimensions in higher
education learning, which are: learner language,
academic discourse, and language in the real-world.
Managing these dimensions are the core for successful
language pedagogy. Aligning learner language,
academic excellence, and resonance to real-world needs
are ideal portrayal of a successful learning. However,
when the ideal is not achieved, identification of the
contributing factor needs to be taken carefully. In
language pedagogy failing a course does not necessarily
mean under performance. Differences in learner
language may be caused by other factors, such as
different social and cultural experiences.

Following the model, EDM needs to consider the
attributes in the learner language, academic discourse,
and real-world language as mentioned below:
4.1 Learner language

There are two sub-dimensions for learner, which are learner’s academic success and learner’s social success.

4.1.1 Academic experience

Attributes for learner’s academic experience are: (1) allotted learning time, (2) score, attempts, and progress made by learner, (3) on-time completion of the tasks, assignments, and course, (4) pause in learning, and (5) personal development plan. Data taken from learners on the amount of time used to access, read, and discussed learning materials would be important signals for interest and commitment in learning. It also indicates learner’s belief in the knowledge they are exposed to. Other attributes could also predict the patterns of learner’s academic routine. These include: the attempts made to do exercises, quizzes, and questions. Attempts made could also include the times learners re-take courses. These may be caused by several reasons: the course is a pre-requisite to other course, the learners failed the course due to non-academic reasons, or learners retake courses to achieve higher marks. In language pedagogy, there is no finite answer to communication problems or language performance. Therefore, retaking course as an attribute needs to be taken as a complement to score. Pause and on-time completion in learning journey are also data to predict learning directions.

Learner may also be asked about the study plan. The plan could predict self-regulation or ownership of the experience in academic context.
4.1.2 Social experience

Attributes for learner’s social experience are: (1) social activities, (2) social media posts, (3) well-being, (4) language attainment could be predicted by their social activities and contributions. Social media activities also indicate of openness and institutions need to be prepared with learners accessing materials outside the textbooks and references provided by the university [29, 30]. Such provision would indicate whether learning environment in the higher education institution was not the source of hindrance to learning. Technology support is also an indication of openness and institutions need to be prepared with learners accessing materials outside the textbooks and references provided by the university [31-33]. Acknowledging different modes in learning is also an attribute as learners have different learning styles and learning preferences [34, 35].

With the provision of genres, language pedagogy is central in supporting learners to answer their needs and to achieve their best in learning. The attribute to guarantee quality in pedagogy is through research and publication [35]. Learners and academics’ participation in research and publication could be used to predict continuous pursuit of excellence in learning.

4.2 Language in academic discourse

The attributes to describe the use of language in the academic discourse are: (1) varieties of genre, (2) technology support, (4) research and publication. Outside the learners’ environment, prediction of learning success may be drawn by looking at the quality of the environment [32]. Provisions for learning success in higher education level are given far beyond lecturing, tutoring, and administrative support. The provision of language environment in support of academic discourse could also be predictors of learning success [26, 27, 28]. Language programs need to integrate varieties of genre to the syllabus, as well as opening access to corpus [29, 30]. Such provision would indicate whether learning environment in the higher education institution was not the source of hindrance to learning. Technology support is also an indication of openness and institutions need to be prepared with learners accessing materials outside the textbooks and references provided by the university [31-33]. Acknowledging different modes in learning is also an attribute as learners have different learning styles and learning preferences [34, 35].

With the provision of genres, language pedagogy is central in supporting learners to answer their needs and to achieve their best in learning. The attribute to guarantee quality in pedagogy is through research and publication [35]. Learners and academics’ participation in research and publication could be used to predict continuous pursuit of excellence in learning.

4.3 Language in the real world

Attributes in this dimension are: (1) reference corpus exposure, (2) industrial experience or internship, (3) community projects, and (4) external evaluation. Prediction of learning and successful language pedagogy could also be drawn from how relevant is learning with the skills needed in the real world. Provision and use of reference corpus in higher education level means provision of natural environment where the language is needed. Reference corpus could include the provision of texts from various media, documents, written and in audio-visual formats. Industrial and professional experiences attained through internship also guaranteed exposure to language skills needed in the real world [36]. When this attribute is found in a language learner or in a language course, it could be marked as an indicator of learning success. Learners could naturally be exposed to the language and language course is informed with materials most needed in the curriculum [37]. Likewise, engaging in community projects not only provides enough opportunity for learners to practice the language but also for an indicator for language course that the skills taught in the institution could answer real-life problems [38].

Finally, external evaluator is considered as an attribute for language pedagogy in higher education. The reason being evaluation in language performance should not be done by several experts within the community of knowledge. Most importantly, evaluation of learning needs to be done by external parties, including professionals and people from industries.

The EDM model would not only provide snapshot of language learners development but also to predict in which direction the learning would take place [39]. To answer the gap in the previous studies, most EDM models focus on the past experiences of learners and predicts the success based on what already happened. This present study offered an EDM model, in which potentials of learners could also be identified through the data gathered in the attributes. The outcome would be a learner’s potential academic routes to be taken by the learner. This model re-defines EDM not as a tool of confirmation but as a true prediction system.

4.4 Business applicability of the model

A pedagogical model was presented by this present study. Its initial aim was to provide a powerful EDM based on the philosophy of co-construction of knowledge. Higher education institutions should not be factories reproducing old knowledge. On the contrary higher education has its full capacity to create new knowledge whilst instilling core values during the process. The EDM model also beneficial to be implemented for business ventures. By predicting what is needed in the industry an institution could evolve its program and make the necessary developments. It could also be used to tap into the changing needs of different generations of learners. Through constantly innovating and re-inventing itself a higher education institution could provide the best value proposition to the market and retain its high reputation in the society.
5 Conclusion

This present study has discussed the importance of EDM for language pedagogy. The present study used and developed the EDM for a dynamic skill, which is skills of using language in many aspects of life. This EDM is not only be useful to predict learners’ success it could also provide early warning system for the educators and learners. Also, program developers could use this model to keep in line with the needs of industry and learners. For practical implications, EDM results would benefit curriculum developers and instructors to design relevant and most current language contents for the teaching and training purposes. Higher education institutions could also use the EDM for assessing their courses and programs and to create new programs for better financial assurance. With healthy and developing ventures language pedagogy would offer best learning experience for learners and impact for the society. Implications for further research would be to apply this EDM to various higher educational institutions to develop its model through contrastive analysis. The research could also be conducted to assess online and onsite learning environment, providing more options post-COVID-19-pandemic time.

This present research is made possible by the sincere gratitude for the continuous support from Group Digital Language and Behavior (RIG D-LAB) at BINUS university.

References

1. P. Bachhal, Jour. of Phys. 12 (2021)
24. D. J. Salas, S. Baldiris, R. Fabregat, S. Graf, Supporting the Acquisition of Scientific Skills by the Use of Learning Analytics, in Int. Conf. on Web-Bas. Lea. (2016)
28. H. Jin, Y. Karatay, F. Bordbarjavidi, J. Yang, T. Kochem, A. A. Muhammad, V. Hegelheimer, ReCALL 34 (2022)
34. C. Bazerman et. al. (Eds.). Traditions of writing research (Routledge, 2010)
35. J. M. Swales, Genre analysis: English in academic and research settings (CUP, 1990)
