

Learning Analytics-Based Approaches for Predicting English Learning Performance

Fatma Yasemin Bayındır Özkan^{1*}

^{1*}Ankara Hacı Bayram Veli Üniversitesi, Turkey. E-mail: fatmayaseminbayindir@gmail.com,
Orcid: <https://orcid.org/0000-0002-8246-4743>

Abstract: The digital ELearning environment provides for continuous streams of behavioral and performance data, thus offering opportunities for LA to assist in detecting learner needs where traditional assessments at the end of a term come too late. This paper presents a synthesis of empirical evidence from institutional, higher education, secondary school, and MOOC settings to analyze how LA and EDM can be used to predict the performance of learners in English. The paper examines three types of prediction methods: procedural and behavioral modeling focusing on sequence and not only the frequency of behaviors; new multimodal data sources in the form of wearable sensors providing information about the learner's cognitive load and anxiety while listening and speaking; and real-time dashboards and ranking systems that transform the predictions in classroom-facing feedback. The general LA methods are discussed with regard to specific English settings, such as the academic language gaps among college transfer students, the experiment with flipped-classroom feedback and the use of serious games to practice vocabulary and pronunciation. Based on these findings, the article suggests a theoretical model for developing LA systems for English learners consisting of four layers - data, analytics, pedagogy, and equity - stressing the need for interpretation of the predictions rather than delivering unmediated judgements to the learner. The conclusions made by the article include the following – LA technology is promising for the early identification of at-risk English learners and for customizing instructions to them. However, the value of LA is contingent upon the continued focus on transparency, learner empowerment and equity especially for multilingual and non-conventional learners for whom there is little data in the training set.

Key Words: learning analytics, english language learning, predictive modelling, learner performance, educational data mining

(Received: 16 March 2024; Revised: 26 April 2024; Accepted: 31 May 2024; Published: 28 June 2024)

Introduction

Teaching and learning of English as a second language, a foreign language, or an additional language have greatly changed with the use of learning management systems (LMS), online courses, and mobile apps. Each time a learner clicks, submits work, posts on discussions, and attempts to take quizzes in such platforms, he/she leaves a digital footprint. These footprints combined provide a wealth of information which has not yet been fully utilized to gain understanding of English proficiency development. Traditional measures used to assess English learning performance like examinations at the end of terms, essays, and oral interviews are helpful to measure learners' achievements and inform tutors of learners' progress but are too late in warning tutors of potential problems. They focus on results rather than on processes. Learning analytics (LA) are defined as the measures, collection, analysis, and reporting of data about learners for purposes of understanding and optimizing the learning process.

Learning analytics is an area of research that has received growing attention over the last decade, with bibliometric analysis revealing its continued growth over time (Chen et al., 2022). Universities and colleges

have started using learning analytics not just for supporting learners on an individual level but also to assess courses and improve their quality based on the information about learners collected through the course of studying (Armatas et al., 2022). In spite of all the developments in this realm, there is still relatively little use of learning analytics to study English language learning in particular, despite the fact that the field of language learning produces information that is richer and more diverse than other disciplines.

This paper fills this void by exploring how learning analytics-based models, especially predictive models, can be utilized for predicting and improving English language learning performance. It compiles insights from educational data mining and learning analytics literature; from studies on MOOCs, flipped classes, serious gaming, and wearable device analytics; as well as from existing research that specifically targets language learning needs. The paper is structured as follows. Section 2 presents the evolution and the scope of learning analytics in education. Section 3 provides an overview of predictive modelling approaches, including behaviour- and procedure-driven prediction, emerging multimodal data sources, and real-time dashboard prediction systems. Section 4 highlights the use of predictive modelling approaches in the specific case of English language learning. Section 5 formulates an integrated theoretical framework for predicting English language learning performance. Sections 6 deals with ethical, equity, and transparency issues, while Section 7 covers the challenges and future directions before concluding in Section 8.

The Evolution and Scope of Learning Analytics in Education

While there are many similarities between learning analytics and educational data mining (EDM), the latter focuses more on pattern detection in large educational datasets, while the former focuses more on the application of such patterns in educational settings. A systematic review of blended learning environments indicates that both disciplines have converged considerably in practical settings because blended courses provide the mixed dataset that both EDM and LA techniques can benefit from. This convergence has led to the development of new techniques to monitor the learners' progress in blended learning environments (Nuankaew et al., 2023).

The bibliographic map of the learning analytics field within the last ten years indicates that research topics have evolved from definitional to applications involving prediction, dashboard development, and engagement analysis as the field develops from concept verification studies to practical implementation (Chen et al., 2022). In a review of the application of learning analytics to improve learner performance in learning portals, various analytics tools used were classified into descriptive analytics, where they describe what has already happened; predictive analytics, where they predict future results; and prescriptive analytics, where they give suggestions for improvement, stating that the use of predictive methods is the most rapidly growing category (Shabnam Ara et al., 2023).

The use of learning analytics has now moved from the realm of higher education to secondary education, albeit much more slowly. A review of the literature on the use of learning analytics in high schools revealed that there was significantly less and less coherent literature on the matter when compared to the tertiary context, with the secondary literature mainly involving generic metrics of student engagement and not performance prediction in specific subjects, suggesting an avenue for future research in English language teaching (Sousa et al., 2021).

Predictive Modelling Approaches for Learning Performance

Behavioural and Procedural Data-Driven Prediction

The key technical significance of learning analytics within the field of performance prediction is represented by the development of the model that will transform learner behavior into predictions of the outcome. One particular evaluation model developed for massive open online courses was based on procedural behavior – the sequence of actions made by a learner rather than just their numbers – and created algorithms that were able to predict the learning performance better than other models based on end-of-course static data (Tong & Zhan, 2023). This procedural focus is especially important for English language learning since language skills develop procedurally; the way a learner practices vocabulary, grammar, and production might be as predictive as the amount of practice.

Prediction models have also been used to identify students who are at risk of poor performance or dropping out. In a study conducted in an online flipped anatomy class, the use of learning analytics methods on LMS logs was able to identify students who were at risk and helped the instructor to provide intervention before the summative assessment based on features like login rate, video watching completion rate and quizzes attempts (Bayazit et al., 2022). While the prediction model was not designed specifically for language classrooms, its rationale – which is identifying learners at risk before a summative test reveals any problem – is easily transposable to English language classes.

Emerging Data Sources: Wearable and Multimodal Analytics

Besides traditional clickstream data from LMS logs, a systematic review of a decade of research literature on learning analytics through wearables found an increasing number of studies that use physiological and behavioural signals, such as eye tracking, heart rate variability, and movements, to measure the aspects of learning like attention, cognitive load, and stress that cannot be captured through clickstream data alone (Liu et al., 2022). Since there are some English learning tasks that have inherent multimodality, such as listening comprehension tasks and oral interactions, physiological indicators seem to offer an exciting but developing approach to complement textual and interactional analyses and enable predictive models to take into consideration the cognitive load in listening tasks and anxiety in speaking evaluations.

Real-Time Dashboards and Performance-Ranking Systems

Predictive analytic solutions have also been adapted for use in operational systems. For instance, a real-time learning analytics and performance ranking forum that enables learners to track their progress and position compared to other learners in terms of learning progress is one such application that integrates the predictive analytic computations with a game-based interface to retain learners' interest (Yi et al., 2022). However, research on the effect of such interfaces is contradictory: research done on dashboard usage in higher education revealed that provision of access to a learning analytics dashboard led to small gains in motivation and self-regulation but strongly relied on the integration of the dashboard into the course structure as well as guidance in interpreting the information from the dashboard (Aguilar et al., 2021). It is an important consideration in terms of applications for English language learning, which suggests that without sufficient pedagogic support, such a dashboard will be useless.

Applying Learning Analytics to English Language Learning Contexts

Identifying Language-Specific Learning Needs

Very few studies have tried using learning analytics in detecting language needs rather than academic performance in general. The studies involving college transfers have found specific deficiencies in academic English language skills, such as deficiencies with subject-specific vocabulary and academic writing conventions, that could not be uncovered with traditional placement test scores and claimed that the data-driven detection would ensure a more equal opportunity transition process for college transfers who are often underserved by the same approach to English language help (Foung et al., 2023). The example proves that learning analytics is capable of going beyond the measurement of engagement and finding specific linguistic constructs, whether lexical, grammatical, or rhetorical, that predict English language performance.

Engagement, Feedback, and Motivation in Skill Development

Feedback based on learning analytics has been experimentally validated in flipped classroom settings in higher education, where students were provided with automatically generated feedback on which pre-class materials they needed to review prior to participating in in-class learning activities assisted by analytics; the experimental group that received the feedback performed better than the control group in post-assessment tasks, indicating that feedback generated by learning analytics can result in improved learning results when it is incorporated into the active pedagogical approach of a flipped classroom (Ustun et al., 2023). This model easily applies to English teaching because flipped classrooms are widely used in this context for grammar and vocabulary lessons.

An analysis of the learning analytics literature focusing on student engagement identified time-on-task, the frequency of participation in discussion forums, and the regularity of assignment submission as reliable performance outcome predictors in various subjects, while also proposing that engagement metrics be integrated into the analysis of content-specific performance metrics (Johar et al., 2023). In the context of English courses based extensively on discussion forums and peer feedback sessions, along with writing assignments, these engagement metrics provide accessible input that can be utilized in performance prediction models.

Serious Games and Interactive Practice Tools

Game-based techniques for vocabulary and grammar training in English language learning are popular, and a systematic review of the use of learning analytics in serious games indicates that integrating learning analytics into the game design reveals usability issues and errors in game design as well as making it possible to adjust difficulty levels in accordance with user performance (Daoudi, 2022). This opens up an interesting opportunity for the implementation of the proposed approach: through the integration of learning analytics in pronunciation and vocabulary games, both better tools and data will be provided.

Younger Learners and Secondary Education

As mentioned in section two, the use of learning analytics in secondary schools is not well documented (Sousa et al., 2021). More importantly, there is little work done in English language teaching of younger students. Considering the fact that the earlier language difficulties are detected, the better it is since more time can be allowed to solve any issue raised, this could possibly be one of the best areas for future work mentioned in section seven.

A Proposed Conceptual Framework for Predicting English Learning Performance

In light of the various types of evidence discussed above, this paper presents a four-tier conceptual model for designing learning analytics systems to predict and promote English learning performance. In the data layer, the evidence is gathered from LMS interaction logs, procedural behavior logs, discussion forums, as well as, wherever possible, multi-modal and/or wearables-based signals relevant to listening and speaking activities. In the analytics layer, the evidence is used in combination with predictive modeling techniques, based on procedural and behavioral approaches, to predict early performance and risk. In the pedagogical layer, the predictions inform teaching through feedback through dashboards, flipped-classroom advice, and serious game difficulty settings, always mediated through teachers' interpretations and not presented directly to the learner as unmediated judgements.

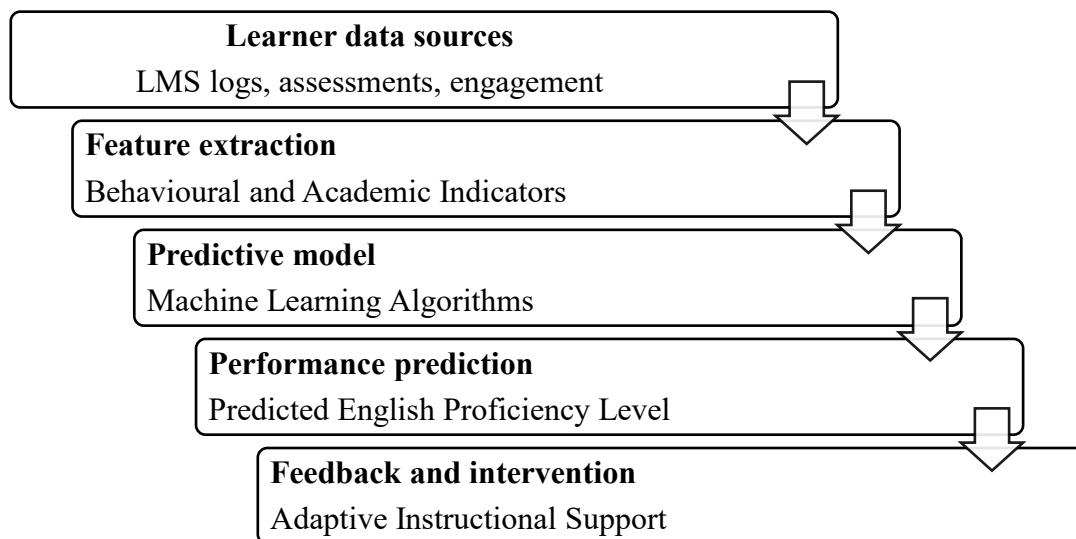


Figure 1: Conceptual Framework for Predicting English Learning Performance

Figure 1 above is a schematic representation of the pipeline with five stages used for going from the learner's actions to the instruction.

Table 1: Representative Predictor Categories for English Learning Performance Drawn from the Learning Analytics Literature

Predictor Category	Representative Data Source	Supporting Studies
Procedural / sequential behaviour	Order and timing of LMS actions, clickstream sequences	Tong & Zhan (2023)
Early risk indicators	Login frequency, video completion, quiz attempts	Bayazit et al. (2022)
Physiological / multimodal signals	Eye-tracking, heart-rate variability, wearable sensors	Liu et al. (2022)
Engagement metrics	Time-on-task, forum participation, submission consistency	Johar et al. (2023)
Language-specific needs indicators	Vocabulary gaps, academic writing performance	Foung et al. (2023)
Game-based interaction data	In-game accuracy, response time, difficulty progression	Daoudi (2022)

The data about the learner (LMS logging, test results, and learner's activity records) is collected in the first stage. The second stage is the process of extracting features from this data that will help create indicators of behaviour and achievement. They are passed to the predictive model in the third stage where the performance prediction (level of the learner's English skills) is made using machine learning techniques.

Finally, the equity layer, elaborated in Section 6, ensures that predictions are used transparently and do not put multilingual, transfer, or non-traditional learners at a disadvantage. Table 1 summarizes the predictor categories found in the reviewed literature.

Ethical, Equity, and Transparency Considerations

Predictive analytics in the educational domain presents many valid worries concerning learner agency, equity, and transparency. As described in a commonly referenced study of personalized learning based on learning analytics, the application of predictive analytics can reinforce existing inequalities if the training data includes biased or partial representation of past achievements and if the learners lack visibility of prediction processes and applications, which compromises their agency regarding their learning journey (Tsai et al., 2020). This worry can be applied to the English learning domain, as the application of predictive models that have been trained using data from native or majority populations can put multilingual, English as an additional language learners, or students from other disadvantaged educational backgrounds at a disadvantage compared to students who have attended institutions with more common curricula.

From the institutional perspective, the responsible implementation of learning analytics in programme reviews and quality improvement should ensure that the predictive findings are used not for classification and gatekeeping but also to take action and provide support, in which the governance framework is evident in how findings can be translated into tangible changes for the benefit of the learner (Armatas et al., 2022). When this approach is applied to the English language learning context, it implies that the predictive outputs need to be presented to teachers and learners in such a way that they are considered diagnostic and formative.

Challenges and Future Directions

A number of issues need to be addressed before prediction using learning analytics can be successfully employed at scale for English language learning. Firstly, a lack of validated outcome labels specific to the language stands out; most of the reviewed studies used general academic performance or success in completing the course instead of validated English proficiency progress metrics. Secondly, while interoperability issues between general learning management systems and English-specific educational resources, such as pronunciation or writing feedback systems, persist, it reduces the amount of information that may be used to develop a complete predictive model. Finally, the issues of explainability described above need to be resolved in order to provide valid predictions for critical decision-making (Tsai et al., 2020).

In terms of future work, several directions look promising. First, longitudinal research studying the trajectory of growth of English skills over time through the lens of learning analytics, rather than single course, may help validate the generalisability of the early procedural indicators mentioned above. Second, an extension of wearable/multimodal learning analytics to listening and speaking assessment in particular could address the ongoing challenge in text-centric learning analytics. Considering that there is still a lack of research on learning analytics in secondary schools, studies applying learning analytics to predict the

development of younger English learners could also be of much value (Sousa et al., 2021). Last but not least, with learning analytics moving towards becoming increasingly diversified in the coming years, regular bibliometric review may prove to be helpful in identifying whether there appears to be a separate sub-discipline of language learning analytics developing (Chen et al., 2022).

Conclusion

In this paper, approaches based on learning analytics to predict English language learning performance have been investigated by referring to the existing literature in general learning analytics, predictive models based on process and behavioral data, novel multimodal datasets, and literature on English and other languages specifically. Based on this evidence, it has become clear that the use of predictive learning analytics can greatly help in detecting at-risk English language learners, generating data-driven feedback in flipped and blended English classrooms, as well as improving the usability and adaptivity of gamification-based language practice applications. However, there is no denying the fact that prediction is not enough as dashboards and rankings help in achieving better results only under supportive pedagogical conditions, while predictive models pose real risks of exacerbating inequities when not designed and governed appropriately. The conceptual framework outlined in this paper with its data, analytical, pedagogical, and equitable layers is one possible beginning point for the responsible adoption of learning analytics at an institutional or English language program level. Further research that confirms language-specific indicators of success, applies analytics to secondary students who lack access, and keeps analytics transparent for all students will be critical in realizing the potential of learning analytics within English language education.

References

- Aguilar, S. J., Karabenick, S. A., Teasley, S. D., & Baek, C. (2021). Associations between learning analytics dashboard exposure and motivation and self-regulated learning. *Computers & Education*, 162, 104085. <https://doi.org/10.1016/j.compedu.2020.104085>
- Armatas, C., Kwong, T., Chun, C., Spratt, C., Chan, D., & Kwan, J. (2022). Learning analytics for programme review: Evidence, analysis, and action to improve student learning outcomes. *Technology, Knowledge and Learning*, 27(2), 461-478. <https://doi.org/10.1007/s10758-021-09559-6>
- Bayazit, A., Apaydin, N., & Gonullu, I. (2022). Predicting at-risk students in an online flipped anatomy course using learning analytics. *Education Sciences*, 12(9), 581. <https://doi.org/10.3390/educsci12090581>
- Chen, X., Zou, D., & Xie, H. (2022). A decade of learning analytics: Structural topic modeling based bibliometric analysis. *Education and Information Technologies*, 27(8), 10517-10561. <https://doi.org/10.1007/s10639-022-11046-z>
- Daoudi, I. (2022). Learning analytics for enhancing the usability of serious games in formal education: A systematic literature review and research agenda. *Education and Information Technologies*, 27(8), 11237-11266. <https://doi.org/10.1007/s10639-022-11087-4>
- Foung, D., Chen, J., & Cheung, K. (2023). Exploring language needs of college transfer students with learning analytics: towards a more equitable experience. *International Journal of Educational Technology in Higher Education*, 20(1), 60. <https://doi.org/10.1186/s41239-023-00429-y>
- Johar, N. A., Kew, S. N., Tasir, Z., & Koh, E. (2023). Learning analytics on student engagement to enhance students' learning performance: A systematic review. *Sustainability*, 15(10), 7849.

<https://doi.org/10.3390/su15107849>

- Liu, Z., Ren, Y., Kong, X., & Liu, S. (2022). Learning analytics based on wearable devices: A systematic literature review from 2011 to 2021. *Journal of Educational Computing Research*, 60(6), 1514-1557. <https://doi.org/10.1177/07356331211064780>
- Nuankaew, P., Jeefoo, P., Nasa-Ngium, P., & Nuankaew, W. S. (2023). Hybrid learning and blended learning in the perspective of educational data mining and learning analytics: A systematic literature reviews. *International Journal of Engineering Trends and Technology*, 71(10), 115-132. <https://doi.org/10.14445/22315381/ijett-v71i10p211>
- Shabnam Ara, S. J., Tanuja, R., Manjula, S. H., & Venugopal, K. R. (2023). A comprehensive survey on usage of learning analytics for enhancing learner's performance in learning portals. *Journal of Educational Technology Systems*, 52(2), 245-273. <https://doi.org/10.1177/00472395231185843>
- Sousa, E. B. D., Alexandre, B., Ferreira Mello, R., Pontual Falcão, T., Vesin, B., & Gašević, D. (2021). Applications of learning analytics in high schools: A systematic literature review. *Frontiers in Artificial Intelligence*, 4, 737891. <https://doi.org/10.3389/frai.2021.737891>
- Tong, Y., & Zhan, Z. (2023). An evaluation model based on procedural behaviors for predicting MOOC learning performance: Students' online learning behavior analytics and algorithms construction. *Interactive Technology and Smart Education*, 20(3), 291-312. <https://doi.org/10.1108/itse-10-2022-0133>
- Tsai, Y. S., Perrotta, C., & Gašević, D. (2020). Empowering learners with personalised learning approaches? Agency, equity and transparency in the context of learning analytics. *Assessment & Evaluation in Higher Education*, 45(4), 554-567. <https://doi.org/10.1080/02602938.2019.1676396>
- Ustun, A. B., Zhang, K., Karaoğlu-Yilmaz, F. G., & Yilmaz, R. (2023). Learning analytics-based feedback and recommendations in flipped classrooms: An experimental study in higher education. *Journal of Research on Technology in Education*, 55(5), 841-857. <https://doi.org/10.1080/15391523.2022.2040401>
- Yi, A. C. Y., Ying, T. K., Yee, S. J., Chin, W. M., & Tin, T. T. (2022). InPath forum: A real-time learning analytics and performance ranking forum system. *IEEE Access*, 10, 128536-128542. <https://doi.org/10.1109/access.2022.3227430>