Tangible Progress: Employing Visual Metaphors and Physical Interfaces in AI-based English Language Learning

Mei Wang
School of Advanced Technology,
Xi'an Jiaotong-Liverpool University
Suzhou, China
Mei.Wang21@student.xjtlu.edu.cn

Hai-Ning Liang
Computational Media and Arts Thrust, Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
hainingliang@hkust-gz.edu.cn

Yu Liu
School of Advanced Technology,
Xi'an Jiaotong-Liverpool University
Suzhou, China
Yu.Liu02@xjtlu.edu.cn

Chengtao Ji
School of Advanced Technology,
Xi'an Jiaotong-Liverpool University
Suzhou, China
Chengtao.Ji@xjtlu.edu.cn

Lingyun Yu
School of Advanced Technology,
Xi'an Jiaotong-Liverpool University
Suzhou, China
Lingyun.Yu@xjtlu.edu.cn

ABSTRACT

Over the past decade, Artificial Intelligence (AI) has made significant strides across various domains, including education. A prominent example of this is the integration of AI-driven language learning tools featuring Automated Essay Scoring (AES) systems. Traditionally, AES relied on predefined criteria, providing scores in simple text formats. In order to enhance learnability and engage learners, we propose a system that harnesses AI-powered AES with a visualization approach. This system consists of three main components: an AI-driven scoring algorithm, a visualization interface translating scoring outcomes into visual metaphors, and tangible posters for presenting scores. We evaluated this visualization system, along with the tangible-formatted results, through domain expert interviews and a three-stage user study. The results indicate that the progressive visual feedback and tangible postcards increased practice frequency and significantly boosted study motivation. Tangible visual feedback demonstrates positive effects on progressive learning.

VLDB Workshop Reference Format:
Mei Wang, Hai-Ning Liang, Yu Liu, Chengtao Ji, and Lingyun Yu. Tangible Progress: Employing Visual Metaphors and Physical Interfaces in AI-based English Language Learning. VLDB 2024 Workshop: Relational Models.

VLDB Workshop Artifact Availability:
The source code, data, and/or other artifacts have been made available at empty.

1 INTRODUCTION

Artificial Intelligence has revolutionized language learning, particularly in the domain of automated essay scoring (AES) [26]. By employing deep learning-based [19] scoring algorithms including advanced Transformer models [21] and multiple types of neural networks, AI-powered AES systems have significantly enhanced the accuracy and efficiency of essay evaluations and have been widely used in recent English language learning practice [15]. For example, "e-rater®" [5] automated essay scoring engine has now replaced human scoring in TOEFL essay writing assessments. Despite these advancements, traditional AI-based AES interfaces typically present learners with raw numerical scores and text-based feedback. This presentation style and format often fall short in several critical aspects. Firstly, there is often a misalignment between the scores and marking dimensions. Traditional systems automatically assess essays and provide final scores, but they do not offer a clear explanation of how these scores are determined based on specific marking criteria (e.g. grammar, coherence, length) or how the essay aligns with the given topic. This lack of clarity in the evaluation process can hinder effective comprehension and impede learners' ability to identify areas for improvement [3]. Secondly, these systems often lack user interactions with the interface. Interactivity, defined as the degree of user involvement with information content and the interface [33], is crucial for fostering engagement and understanding. An interactive interface can shape users' perceptions of AI products [35]. However, current AI-based AES interfaces often present learners with a static, one-way communication model, where they can only passively view their scores without opportunities for active interaction or exploration in their learning progress. Thirdly, feedback is essential in language learning as it guides improvement [16] and fosters motivation [36], which is vital to learners' educational experience. However, traditional UI designs in AES systems offer limited encouragement through their feedback mechanism. These systems mainly rely on numerical scores and simple elements like star ratings to denote performance levels. While these methods can convey accurate results, they lack visual appeal and engagement. This shortfall in design fails to provide an aesthetically pleasing experience that could potentially enhance the learners' motivation and engagement.

To overcome these limitations, we have developed an innovative system that integrates visual metaphors with AI-powered essay scoring technology, enhancing interactivity and user-centered

Proceedings of the VLDB Endowment. ISSN 2150-8097.

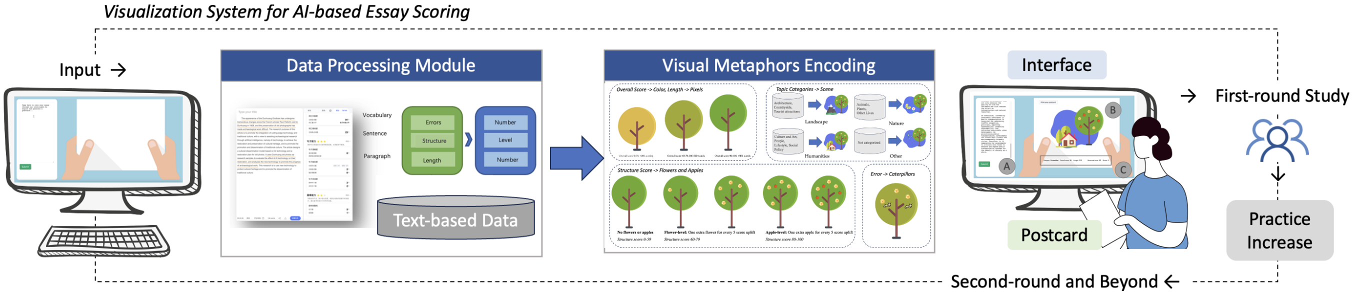


Figure 1: The framework of the AI-based visual system for English language learning includes the scoring algorithm, visual feedback, and postcards for presenting scores. A three-round study was conducted to evaluate its usability in progressive language learning.

learning experience through tangible-based interface (Fig 1). Our system incorporates an improved algorithm that leverages seq2seq models based on Transformer and RNN [30] to generate score dimensions. To enhance learners' engagement in the progressive learning and facilitate the identification of writing shortcomings more effectively, we designed an interactive system with visual representations. This system visually displays scores based on individual marking criteria, along with metrics on the learners' performance and progress. Furthermore, to intuitively demonstrate and motivate learners by illustrating their progress, our system presents visual feedback of scores in a postcard format. Visual changes on the postcards reflect improvements and developments in learners skills. This approach not only provides efficient feedback but also enhances learners' comprehension and self-reflection on their essay-writing progression, making the learning process both engaging and informative.

To evaluate the efficacy of our prototype in language education, we engaged in expert interviews and conducted a series of progressive user studies. These studies were aimed at assessing the design and usability of our visualization prototype and delving into the user experience and learning effects, particularly concerning the tangible format and its influence on progressive learning.

2 RELATED WORK

In this section, we provide an overview of Automated Essay Scoring and interactive visual systems for English language learning.

Automated Essay Scoring. Different from the traditional teaching methodologies, recent developments in English education have shifted towards technology-based solutions, with digital approaches significantly transforming pedagogy [11]. Among these, AES stands out as a well-established and rapidly evolving technology that greatly benefits English language learning. AES is a computer-based assessment approach that automatically scores essays by analyzing various features. Early AES techniques [26] employed simple rule-based algorithms, whereas modern AI-based techniques use machine learning and NLP to offer more accurate scoring. Modern methods include syntactic and semantic analysis, text classification, and sentiment analysis, which allow for a more comprehensive assessment of essays [5]. Furthermore, machine learning algorithms such as support vector machines and neural networks [30] are commonly used for classification and scoring tasks, enabling AES systems to provide automated evaluation and achieve state-of-the-art performance [31]. With the recent advancements in language models [28], the Transformer [21] has emerged as a leading approach in AES. In this study, we use the state-of-the-art seq2seq models based on Transformer and LSTM [30] to develop an improved AES algorithm which generates scores that are used for further processing.

Interactive System for English Language Learning Visualization can enhance people's understanding and retention of information by leveraging techniques that improve visual perception. As a means to efficiently present abstract information [2], visualization becomes particularly useful in the context of English language learning, where it can assist learners in better grasping abstract language concepts. Visual metaphors emerge as a powerful form of visualization. They serve as an effective tool for conveying complex ideas, enhancing message clarity and emotional engagement. Current studies [18] explore how these metaphors can convey complex ideas through the representation of incongruity in visual elements. Furthermore, Ventalon's work [37] delves into the cognitive processes that underpin the understanding of these metaphors. Additionally, many visualization strategies have been examined to enhance user comprehension and introspection of their personal data [7]. Differing from conventional task-centric information visualization systems, casual visualization is tailored for daily usage [25]. This casual visualization typically uses pleasing aesthetic elements to encourage non-critical awareness and persistent behavioral modifications [22]. For instance, UbiFit Garden [8] employs a blossoming garden metaphor to denote user's physical exertions. Likewise, FishnSteps [20] establishes a correlation between a user's daily step tally and the animation of a fish in a digital environment. Sun et al. [32] transformed food posts on social media into postcards and also proposed butterflies-shaped artistic representation for biofeedback. In this regard, visualization stands out as an important and effective method of enhancing user experience by presenting information graphically and intuitively to learners [24].

Additionally, user interaction is also an important factor in the language education [1]. Interactivity not only enhances user engagement but also facilitates better understanding and communication among learners, thereby enriching their overall educational experience [14]. Although studies have shown that interactive visualization systems foster stronger emotional connections compared to systems that rely solely on statistical reports [23], the integration of visualization systems into AES systems is still in its early stages. Current AES systems typically generate numerical data and rely on text-based interface, providing limited visual-appeal feedback to learners [29]. Therefore, we aim to innovate within this area by employing aesthetic codes to transform numerical data into visually engaging representations, thereby redefining the conventional format of "scores".

3 SYSTEM DESIGN AND IMPLEMENTATION

In this section, we present the design considerations and goals. Then, we introduce our visual prototype for essay marking.

3.1 Design Goals

To better understand the requirements of language educators, we conducted one-to-one interviews with six domain experts in language teaching (E1-E6, consisting of 3 females and 3 males, with a mean age of 32). Through these discussions, we identified the following design goals:

G1. Provide comprehensive feedback: The system should offer detailed feedback on students' English writing, covering aspects such as *grammar*, *spelling*, *content coherence*, and *thematic clarity*.
G2. Enhance and maintain studying interest: To boost students' interest in learning, the system provides intuitive and engaging feedback to present students' writing proficiency.
G3. Offer tangible representations of progress: Present studying progress in a tangible, organized, and easily understandable format. This helps learners clearly see their achievements and improvements over time, thereby enhancing motivation and engagement.

3.2 The Visualization System for AES

To achieve our goals, we developed an essay writing and marking prototype that includes a scoring algorithm powered by NLP technologies, paired with a visualization interface that translates numeric scores into visual metaphors. Furthermore, we printed the visual feedback on postcards, providing a tangible representation of students' progressive improvements.

3.2.1 Scoring Algorithm. Our system features an AI-based scoring algorithm, which involves two primary modules.
AES Scoring Dimensions. The AES module evaluates essays on four dimensions: Our scoring solution incorporates Microsoft Aimwriting[1], a cutting-edge tool designed to enhance the accuracy and efficiency of essay scoring. We chose to build and improve upon Aimwriting due to its comprehensive framework which is supported by several advanced NLP models, extensive English language resources, and deep expertise in essay evaluation rooted in foundational research. This framework include several dimensions.

- `Structure`:This dimension assesses inter-sentence coherence and fluency. Coherence at the whole essay level is evaluated using lexical chains and LexRank [10], which determine the coherence of discourses or focus on topics within an essay. Organization is analyzed by examining the arrangement of paragraphs, including the introduction, body, and conclusion. Fluency is assessed through an N-gram [6] statistical language model and a bi-directional LSTM neural model [30]. The N-gram model evaluates the likelihood of word combinations within proximity, capturing contextual information, while the LSTM model analyzes word dependency within sentences, rewarding sentences that demonstrate high fluency.
- `Length`:This dimension evaluates the informativeness of an essay. Factors such as the overall length of the essay, the number of paragraphs, and the ratio of high IDF (Inverse Document Frequency) words [27] are analyzed to assess the sufficiency of information provided, which helps determine whether the essay contains an appropriate depth of content .
- `Errors`:The system identifies and tallies spelling and grammar errors using a seq2seq model based on Transformer while enhanced by fluency boost learning [13]. This approach ensures the accurate detection of errors, contributing to a detailed assessment of the writing's correctness.
- `Overall Score`: Features from the aforementioned dimensions are extracted for each essay from the training dataset. These extracted features are then used to construct a Multiple Additive Regression Trees (MART) [12], which is integral in handling complex predictive modeling challenges. Once the model is constructed, it uses the features of an input essay to assign an overall score that reflects the quality of the essay.

Topic Categorization. To identify topics of input essays, we employ Latent Dirichlet Allocation algorithm (LDA) [17]. We first model each topic as a list of words with LDA, then convert it into final categories through rule-based method. The procedures include preprocessing raw text, converting tokenized and lemmatized essay to a bag of words, running LDA with genism by setting number of topics and finally assign a category according to the topics list as shown in Table 1.

Table 1: Category of topics

Category	Topics
Landscape	Architecture, Countryside, Tourist attractions
Humanities	Culture and Art, People, Lifestyle, Social Policy
Nature	Animals, Plants, Other Lives
Other	Not categorized

3.2.2 Visual Metaphor Design. We collaborated with experienced UX designers to create clear and engaging visual representations for our system. We chose for a flat cartoonish style, which resonates well with students, to make the learning experience more approachable and enjoyable. Utilizing copyright-free design materials from Freepik[2], we finalized the visual designs.
The Metaphor of Tree **:** We employed the visual metaphor of a tree to symbolize language learning progress, illustrating students' growth as a thriving grown plant. This metaphor not only

[1]https://aimwriting.mtutor.engkoo.com/

[2]https://www.freepik.com/.

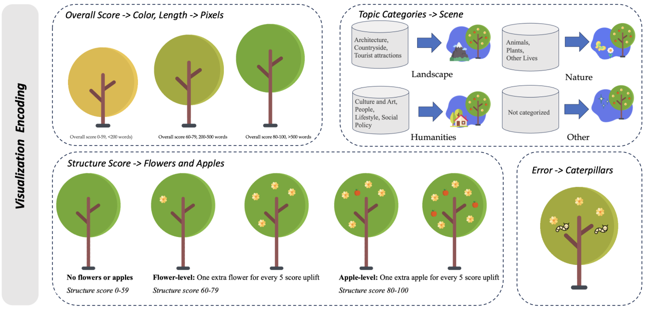


Figure 2: Visual encodings for key scoring dimensions of essay writing: overall score, length, structure, topic relevance, and error count.

enhances positive engagement but also encourages users to nurture their "trees" through consistent learning efforts [9], visually marking their progress and milestones. Studies [34] indicate that metaphorical designs like this can foster greater emotional engagement than purely quantitative representations. Additionally, we transform this design into a tangible format that learners could interact with throughout their learning period, further enhancing the educational experience.

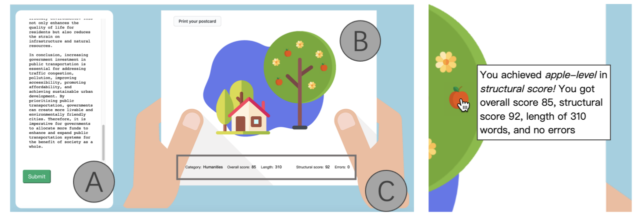
Encoding Rules: Our system visualizes the scores into several ranges, facilitating easy comparison. Like grades, categorizing scores into distinct levels based on predefined ranges is widely used in education [4]. In typical educational settings, scores are often categorized into ranges or bands to assign grades like A, B, C, or F. For example, scores above 80 will normally be considered excellent, while scores below 60 will be considered to fail. Categorizing the scores into different visual representations allows for a clearer understanding of performance levels and facilitates comparisons between different essays or individuals. It also provides a more intuitive and easily communicable representation of the scores in our system. To be specific, we transform scores and the topic categories into visual metaphors as below (Fig. 2):

- `Structural Score`: indicated by the number of flowers and apples, symbolizing different stages of the tree's growth.
- `Length`: represented by the length of the tree trunk.
- `Errors`: demonstrated by the number of caterpillars on the tree.
- `Overall Score`: illustrated by the color of the leaves, where, for instance, a high score corresponds to green leaves, and a low score to yellow leaves.
- `Topics`: a theme-related image displaying as a part of the background besides the tree.

The interface is built with HTML, CSS3, and JavaScript. It allows users to input essays and view visualization results with an optional detailed encoding explanation upon submission (Fig. 3).

4 USER STUDY

We conducted an expert interview with domain experts to discuss the design and teaching approach. Additionally, we carried out a three-stage user study with English language learners to evaluate the effectiveness of our system in the language learning process.



(1) Result page after entering essay (2) Instructions

Figure 3: The visualization results. A: the textbox where users input their essays; B: the visual feedback; C: Original scores. If the user hovers over a visual element, the encoding explanation will appear.

4.1 Domain Expert Evaluation

The domain expert interview was structured to focus on two main aspects: (1) the system's ease of learning and usability, and (2) the effectiveness of the visual metaphor designs in presenting results. We engaged six independent experts three females and three males with diverse backgrounds: two in computer science (E1, E5), two in language teaching (E2, E4), and two in design (E3, E6). We first provided a detailed overview of the system's background and its design principles, including visual encoding examples, to ensure a thorough understanding of the system. Participants then freely explored the system for 30 minutes, after which we collected feedback on its usability, user experience, and potential enhancements.

All the experts (E1-E6) offered positive feedback on our visual encoding system, agreeing that the correlation between the scores and the visual metaphors was clear and easy to understand. They unanimously affirmed that the use of intuitive visual metaphors would enhance the language learning experience for students. Specifically, the language teachers highlighted that this visual system would be particularly beneficial for younger students and expressed their interest in integrating it into their teaching and learning environments. Experts E1 (specializing in computer science) and E3 (specializing in design) were particularly impressed with the use of tangible postcards for presenting results. They highlighted the advantages of a tactile, interactive experience, suggesting that it could lead to greater student engagement and provide a more concrete representation of their achievements. They noted that tangible reports, which can be stored and compared, would offer students a clearer perspective on their learning progress and could motivate them to dedicate more effort to their studies. Additionally, E3 recommended incorporating dynamic effects into the visualization interface, such as rustling leaves, blooming flowers, and twinkling stars, to enhance the appeal and engagement of the reports. E4 (specializing in language teaching) suggested expanding the range of topic categories to encompass broader subjects. Additionally, there were recommendations to include more metaphor themes in future updates to enhance the adaptability of the visual system.

4.2 Three-Stage User Study

The user study followed a structured workflow comprising three rounds of study (Fig. 1), spaced one week apart, aiming to gain a deeper understanding of the system's short-term and potential long-term effects. This study specifically focused on examining the effectiveness of using visual metaphors and the tangible interface to present learning outcomes, targeting university students preparing for the IELTS[3] exam and K12 students learning English language.

Participants. We recruited a group of 22 participants (12 females, 10 males; Mean=20, SD=5.19), including 18 university students aged between 18 to 30), and 4 K12 students aged 10 to 12. The university students' fields of study included computer science, psychology, management, and mathematics. These participants were arranged into two groups in the second round study based on a pair-wise design.

Procedures. At the beginning of the study, participants were provided with written informed consent and then briefed on the purpose of the study. Our first round study aimed to assess user experience with our visualization system for evaluating their essays and providing visual and tangible feedback on scores. For this round, participants were invited to submit essays they had completed in the past week, averaging four essays per individual. These essays were then uploaded to our system, which provided instant visual feedback along with tangible postcards. These postcards featured visual representations and explanations of the visual encoding process. Following this, we collected questionnaires and conducted interviews with the participants to evaluate their experience with the system.

One week later, we conducted the second round of the study, which aimed to evaluate participants' studying motivation and learning outcomes after receiving the results from the first round. We collected six essays from each participant, all written during the interval week. This time, the participants were divided into two groups of 11 each. For the first group ("visual" group), we uploaded their essays and provided the associated tangible postcards with visual representations of their achievements. For the second group ("numeric" group), we assessed their essays using the AES approach but only provided numeric scores. We conducted interviews with the participants to understand their experiences with this progressive learning approach.

One week after the second round, we conducted a third round of the study. Each participant from the two groups uploaded six essays to the system, allowing us to continue assessing their learning progress. Additionally, we conducted interviews to gather further insights into their learning experiences. This third round aimed to evaluate the participants' ongoing study efforts, particularly assessing whether the tangible visual feedback had a positive impact on their progressive learning. We also examined sustained engagement by observing whether the visualization system maintained user engagement over time.

5 RESULTS AND FINDINGS

We conducted a detailed analysis of the findings from our three-round user study, focusing on three key aspects: the outcomes of progressive learning, the impact of visual metaphors on language learning, and the potential of using a tangible form to support progressive learning and sustain engagement.

Progressive Learning. Approximately 86% of questionnaire respondents indicated that the visualization system boosted their motivation to study in the second round of the study, leading to increased practice frequency and expanded coverage of topics. Notably, during the third round, participants in the "visual" group reported engaging in at least two additional practice sessions per week and exploring an average of one additional category after receiving their results from the first and second rounds. This increase was largely motivated by their anticipation of receiving more postcards, highlighting the motivational impact of the tangible feedback provided. While 64% of participants from both groups reported improvements in their study outcomes through the use of the system, we did not observe a significant difference between the two groups in terms of essay writing skill improvement in the third round. This could be attributed to the short duration of the third stage study (one week difference using our system between two groups), which may have limited the assessment of performance improvement. Most of the participants (82%) expressed their willingness to use the visualization system and receive visual tangible feedback of their results progressively in the future. Particularly, participants in the "visual" group showed stronger intentions and were more willing to engage with the system again.

Visual Metaphors. In comparing visual feedback and traditional numeric feedback, 73% of participants expressed a preference for visual feedback. The remaining participants appreciated the visual feedback but indicated that they also needed precise numeric results. This combination helps them gain a clearer understanding of their improvements and assess their progress more accurately. This feedback suggests that while visual feedback is engaging and preferred by the majority, incorporating both visual and numeric feedback could cater to a broader range of learning preferences and enhance the overall effectiveness of the system. Regarding the visual designs, all participants found the visual metaphors intuitive and expressive. They were able to easily understand the mappings and observe their strengths and areas for improvement through the visual presentations. Participants also suggested expanding the scenes that represent different topics. They expressed a desire to collect these visualizations and document their visual progress, using them as a tangible reminder of the extent and depth of their practice. This feedback indicates a strong engagement with the visual aspect of the learning tool, highlighting its potential as a motivational aid in educational settings.

Tangible Feedback. In the first round interview, most participants felt curious and interested upon encountering the postcard format. Other reactions included feelings of surprise, confusion, novelty, and excitement. In subsequent interviews conducted in the second and third round, participants overwhelmingly expressed a desire to collect tangible visual representations of their progress. A significant majority (over 80%) strongly favored the postcard design, particularly appreciating its size and thematic elements. This collection serves not just as a motivational tool but also as a personal record of their educational journey. It provides a tangible reminder of their growth and achievements, which further reinforces the learning process. Although the study period was limited to observe significant improvements in essay writing skills compared to the

[3]International English Language Testing System: https://ielts.org/

traditional approaches, the heightened studying motivation observed suggests that students' learning outcomes could be greatly enhanced with long-term use.

6 CONCLUSION

Throughout our study, we observed that integrating visual and tangible feedback mechanisms potentially enhances continuous learning and improvement. This approach proves particularly valuable in educational settings, especially in areas such as language learning where gradual progress is essential. Currently, we have attempted to use tangible postcards, which are cost-effective and easy to collect. Looking ahead, it would be intriguing to explore other forms of tangible feedback. For instance, implementing a progressive puzzle-solving or building activity could provide a dynamic and interactive way for learners to visualize their progress. This could not only maintain high levels of engagement but also deepen the learners' connection to their educational journey, making the learning process both enjoyable and impactful.

As Large Language Models (LLMs) have revolutionized the interaction with language, we will also explore the future possibility of employing LLMs to AES in the next stage. By incorporating LLMs' advanced text understanding capability, we expect to have a new UI involving more interaction and communication between the user and the interface, which can offer more nuanced feedback, while aligning with educational goals and offering a more holistic evaluation.

REFERENCES

[1] Dr Mohammad Reza Ahmadi. 2018. The use of technology in English language learning: A literature review. *International journal of research in English education* 3, 2 (2018), 115–125. https://doi.org/10.29252/ijree.3.2.115

[2] Nouf Matar Alzahrani. 2020. Augmented reality: A systematic review of its benefits and challenges in e-learning contexts. *Applied Sciences* 10, 16 (2020), 5660. https://doi.org/10.3390/app10165660

[3] Evelin Amorim, Marcia Cançado, and Adriano Veloso. 2018. Automated essay scoring in the presence of biased ratings. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. 229–237. https://doi.org/10.18653/v1/N18-1021

[4] Judith Arter and Jay McTighe. 2001. *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance*. Corwin Press.

[5] Jill Burstein, Joel Tetreault, and Nitin Madnani. 2013. The e-rater® automated essay scoring system. In *Handbook of automated essay evaluation*. Routledge, 77–89.

[6] William B. Cavnar and John M. Trenkle. 1994. N-Gram-Based Text Categorization. In *Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*. Las Vegas, US, 161–175.

[7] E. Choe, B. Lee, and m. schraefel. 2015. Characterizing Visualization Insights from Quantified Selfers' Personal Data Presentations. *IEEE Computer Graphics and Applications* 35, 04 (jul 2015), 28–37. https://doi.org/10.1109/MCG.2015.51

[8] Sunny Consolvo, David W. McDonald, Tammy Toscos, Mike Y. Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, Ian Smith, and James A. Landay. 2008. Activity Sensing in the Wild: A Field Trial of Ubifit Garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 17971806. https://doi.org/10.1145/1357054.1357335

[9] Ed de Quincey, Chris Briggs, Theocharis Kyriacou, and Richard Waller. 2019. Student centred design of a learning analytics system. In *Proceedings of the 9th international conference on learning analytics & knowledge*. 353–362. https://doi.org/10.1145/3303772.3303793

[10] Günes Erkan and Dragomir R Radev. 2004. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of artificial intelligence research* 22 (2004), 457–479. https://doi.org/10.1613/jair.1523

[11] Tira Nur Fitria. 2021. Grammarly as AI-powered English writing assistant: Students alternative for writing English. *Metathesis: Journal of English Language, Literature, and Teaching* 5, 1 (2021), 65–78. https://doi.org/10.31002/metathesis.v5i1.3519

[12] Jerome H Friedman and Jacqueline J Meulman. 2003. Multiple additive regression trees with application in epidemiology. *Statistics in medicine* 22, 9 (2003), 1365–1381. https://doi.org/10.1002/sim.1501

[13] Tao Ge, Furu Wei, and Ming Zhou. 2018. Fluency boost learning and inference for neural grammatical error correction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*. 1055–1065. https://doi.org/10.18653/v1/P18-1097

[14] Ronghuai Huang, J Michael Spector, Junfeng Yang, Ronghuai Huang, J Michael Spector, and Junfeng Yang. 2019. Learner experiences with educational technology. *Educational technology: A primer for the 21st century* (2019), 91–105. https://doi.org/10.1007/978-981-13-6643-7_6

[15] Xinyi Huang, Di Zou, Gary Cheng, Xieling Chen, and Haoran Xie. 2023. Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society* 26, 1 (2023), 112–131. https://doi.org/10.30191/ETS.202301_26(1).0009

[16] Ken Hyland and Fiona Hyland. 2006. Feedback on second language students' writing. *Language teaching* 39, 2 (2006), 83–101. https://doi.org/10.1017/S0261444806003399

[17] Hamed Jelodar, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, and Liang Zhao. 2019. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia tools and applications* 78 (2019), 15169–15211. https://doi.org/10.1007/s11042-018-6894-4

[18] Zoltán Kövecses. 2020. Visual metaphor in extended conceptual metaphor theory. *Cognitive Linguistic Studies* 7, 1 (2020), 13–30. https://doi.org/10.1075/cogls.00047.kov

[19] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436–444. https://doi.org/10.1038/nature14539

[20] James J. Lin, Lena Mamykina, Silvia Lindtner, Gregory Delajoux, and Henry B. Strub. 2006. Fish'n'Steps: Encouraging Physical Activity with an Interactive Computer Game. In *Proceedings of the 8th International Conference on Ubiquitous Computing* (Orange County, CA) *(UbiComp'06)*. Springer-Verlag, Berlin, Heidelberg, 261278. https://doi.org/10.1007/11853565_16

[21] Sabrina Ludwig, Christian Mayer, Christopher Hansen, Kerstin Eilers, and Steffen Brandt. 2021. Automated essay scoring using transformer models. *Psych* 3, 4 (2021), 897–915. https://doi.org/10.3390/psych3040056

[22] Andrew Vande Moere and Helen Purchase. 2011. On the role of design in information visualization. *Information Visualization* 10, 4 (2011), 356–371. https://doi.org/10.1177/1473871611415996

[23] Richard Moran. 1994. The expression of feeling in imagination. *The philosophical review* 103, 1 (1994), 75–106. https://doi.org/10.2307/2185873

[24] Thomas Naps, Stephen Cooper, Boris Koldehofe, Charles Leska, Guido Rößling, Wanda Dann, Ari Korhonen, Lauri Malmi, Jarmo Rantakokko, Rockford J Ross, et al. 2003. Evaluating the educational impact of visualization. *Acm sigcse bulletin* 35, 4 (2003), 124–136. https://doi.org/10.1145/960492.960540

[25] Zachary Pousman, John Stasko, and Michael Mateas. 2007. Casual information visualization: Depictions of data in everyday life. *IEEE transactions on visualization and computer graphics* 13, 6 (2007), 1145–1152. https://doi.org/10.1109/TVCG.2007.70541

[26] Dadi Ramesh and Suresh Kumar Sanampudi. 2022. An Automated Essay Scoring Systems: A Systematic Literature Review. *Artif. Intell. Rev.* 55, 3 (2022), 24952527. https://doi.org/10.1007/s10462-021-10068-2

[27] Stephen Robertson. 2004. Understanding inverse document frequency: on theoretical arguments for IDF. *Journal of documentation* 60, 5 (2004), 503–520. https://doi.org/10.1108/00220410410560582

[28] Pedro Uria Rodriguez, Amir Jafari, and Christopher M Ormerod. 2019. Language models and automated essay scoring. *arXiv preprint arXiv:1909.09482* (2019). https://doi.org/10.48550/arXiv.1909.09482

[29] Rustam Shadiev and Mengke Yang. 2020. Review of studies on technology-enhanced language learning and teaching. *Sustainability* 12, 2 (2020), 524. https://doi.org/10.3390/su12020524

[30] Alex Sherstinsky. 2020. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena* 404 (2020), 132306. https://doi.org/10.1016/j.physd.2019.132306

[31] Jinnie Shin and Mark J Gierl. 2021. More efficient processes for creating automated essay scoring frameworks: A demonstration of two algorithms. *Language Testing* 38, 2 (2021), 247–272. https://doi.org/10.1177/0265532220937

[32] Zhida Sun, Nan Cao, and Xiaojuan Ma. 2017. Attention, Comprehension, Execution: Effects of Different Designs of Biofeedback Display. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 21322139. https://doi.org/10.1145/3027063.3053082

[33] S Shyam Sundar and Jinhee Kim. 2005. Interactivity and persuasion: Influencing attitudes with information and involvement. *Journal of interactive advertising* 5, 2 (2005), 5–18. https://doi.org/10.1080/15252019.2005.10722097

[34] H. Hong T. Kim and B. S. Magerko. 2010. Design requirements for ambient display that supports sustainable lifestyle. In *Proceedings of the 8th ACM Conference*

on Designing Interactive Systems (Aarhus, Denmark) *(DIS '10)*. Association for Computing Machinery, New York, NY, USA, 103112. https://doi.org/10.1145/1858171.1858192

[35] Wei Tao, Shuang Gao, and Yilang Yuan. 2023. Boundary crossing: an experimental study of individual perceptions toward AIGC. *Frontiers in Psychology* 14 (2023), 1185880. https://doi.org/10.3389/fpsyg.2023.1185880

[36] Elizabeth Tricomi and Samantha DePasque. 2016. The role of feedback in learning and motivation. In *Recent developments in neuroscience research on human motivation*. Vol. 19. Emerald Group Publishing Limited, 175–202. https://doi.org/10.1108/S0749-742320160000019015

[37] Geoffrey Ventalon, Grozdana Erjavec, and Charles Tijus. 2023. A review of processing and analysing visual metaphors in psychology. *European Review of Applied Psychology* 73, 4 (2023), 100836. https://doi.org/10.1016/j.erap.2022.100836