Original Article

Available online at
https://jurnal.uinsyahada.ac.id/index.php/Lavoisier/index
LAVOISIER: CHEMISTRY EDUCATION JOURNAL

Vol. 5 (1), 2026

The Effect of EnglishScore Media-Based Deep Learning on TOEFL Structure Scores in the Chemistry English Course

Dimas Ridho[1*], Lenni Khotimah Harahap[2]

[1] Chemistry Education, Medan State University, Indonesia
[2] Chemistry Education, State Islamic University Walisongo Semarang, Indonesia

*Correspondence: dimas@unimed.ac.id

Article History

Received 04 30th 2026
Revised 06 02nd 2026
Accepted 06 04th 2026
Available Online 06 30th 2026

Keywords:
Deep Learning
English for Chemistry
English Score
Structure and Written Expression
Toefl Score

Abstract

Low TOEFL scores among students in the Structure and Written Expression section are often caused by conventional, superficial learning methods. This study aims to analyze the effect of a deep learning approach supported by the EnglishScore media on students' TOEFL scores in the Chemistry English course. The research method used was a quasi-experimental study with a non-equivalent control group design. The research sample consisted of two classes of second-semester Chemistry Department students at the State University of Medan, selected through purposive sampling. Data were collected through pretests and posttests, which were analyzed using descriptive statistics and the independent samples t-test via SPSS software. The results showed that the average posttest score of the experimental class (28.829) was significantly higher than that of the control class (23.229). The results of the hypothesis test showed a significance value of 0.001 ($p < 0.05$), indicating a significant difference in achievement between the two groups. The deep learning approach proved effective in creating a more homogeneous distribution of scores and enhancing students' conceptual understanding through reflective and analytical activities. The integration of the EnglishScore platform played a crucial role in providing an adaptive learning environment and enhancing cognitive engagement. This study highlights the importance of transforming technology-based learning strategies to strengthen students' English language competencies in the field of science.

DOI: https://doi.org/10.24952/lavoisier.v5i1.19662

1. Introduction

Higher education in the era of globalization requires students, including those in non-language programs such as Chemistry, to possess strong English language proficiency as a means of international scientific literacy (Hyland, 2022). The ideal condition for English language learning at the university level is the creation of an academic environment that facilitates deep learning of language structure, particularly in the area of Structure and Written Expression (Trigwell & Prosser, 2020). Through this approach, students are expected not merely to mechanically memorize grammatical rules, but to internalize linguistic logic to support their professional and academic communication skills. The integration of digital technology into this process should serve as a catalyst that enables students to monitor their learning progress independently and adaptively (Bond et al., 2021; Zawacki-Richter et al., 2019).

However, the reality on the ground reveals significant challenges in students' ability to achieve TOEFL scores, particularly in the Chemistry English course. Many students still view the Structure and Written Expression material as the main obstacle due to its abstract and complex nature. A learning approach that tends to be conventional and focused on final outcomes (surface learning) often makes it difficult for students to apply their understanding of grammar in different contexts. Furthermore, reliance on lecture-based methods without the support of relevant interactive media results in low cognitive engagement among students, which ultimately impacts their TOEFL scores, preventing them from reaching the expected minimum standard (Chen & Moradi, 2023; Han & Gardner, 2021; Kormos, 2020). This situation requires immediate attention; preliminary studies indicate that 65% of chemistry students at the Medan State University have not yet reached the standard score of 450. If left unaddressed, this stagnation in grammar proficiency will hinder students' ability to read international chemistry journals, impede the completion of their final theses, and reduce the competitiveness of chemistry graduates in the technology-driven global job and professional markets.

A limitation of this study is the scarcity of research that integrates deep learning approaches with app-based assessment tools, such as EnglishScore, for science students. Most previous studies have focused more on the effectiveness of digital media in general without addressing the depth of students' cognitive processes or have focused only on language majors (Uzorka & Odebiyi, 2025). In fact, the characteristics of chemistry students require a more structured, analytical, and practical approach. The urgency of implementing EnglishScore in this study stems from its role as a self-assessment platform that provides real-time feedback meeting international standards a crucial factor in addressing the limited face-to-face time in chemistry English classes. Consequently, there is an urgent need to explore how the synergy between learning strategies that emphasize critical reflection and digital tools can reduce the variance in student proficiency and improve the efficiency of English language learning in engineering or science higher education settings (Holmes & Tuomi, 2022; Hossain & Munam, 2022; Markauskaite et al., 2022)

DOI : https://doi.org/10.24952/lavoisier.v5i1.19662

Based on these issues, this study aims to analyze the effect of implementing a deep learning approach supported by the EnglishScore platform on improving students' TOEFL scores, particularly in the areas of Structure and Written Expression. Specifically, this study focuses on second-semester students in the Chemistry Department at the University of North Sumatra to examine the extent to which this intervention is effective in creating a more homogeneous and statistically significant score distribution. The results of this study are expected to provide empirical contributions to the development of more innovative English learning models that align with the spirit of Merdeka Belajar in university.

2. Materials and Methods

This study employed a quantitative approach using a quasi-experimental method through a pretest–posttest nonequivalent control group design. This design was used because the research subjects were drawn from classes that had already been formed within the academic system; therefore, individual randomization was not conducted in order to maintain a conducive academic environment. The study involved two groups of subjects: an experimental class that received deep learning instruction supported by the EnglishScore platform, and a control class that received conventional lecture-based instruction.

The study was conducted at the Department of Chemistry, Medan State University, during the even semester of the 2025/2026 academic year. The study population consisted of all second-semester students in the Chemistry Education Program, distributed across 6 classes with a total of 210 students. Sampling was conducted using purposive sampling based on equivalence of initial ability and academic considerations. Based on this technique, the research sample consisted of 38 students in the experimental class and 35 students in the control class, for a total sample size of 73 students.

The data used in this study consists of quantitative data in the form of students' TOEFL scores on the Structure and Written Expression subtests. The data sources for this study are primary data obtained directly from the results of the test administered to students in both the experimental and control groups during the pre-test (before the intervention) and post-test (after the intervention) sessions.

The primary research instrument used was the TOEFL Structure and Written Expression section, officially adapted by Phillips (2001) from the book Longman Preparation Course for the TOEFL Test. This test instrument consists of 40 multiple-choice items divided into two standard subtests: 15 Structure questions (sentence completion) and 25 Written Expression questions (identifying grammatical errors). Before this instrument was applied in the field, the validity of the document was tested through content validity testing by several experts (expert judgment). Subsequently, the instrument's reliability was tested using the Kuder-Richardson (KR-20) formula to ensure the internal consistency of the measurement tool. In addition to the main test instrument, the EnglishScore application was utilized as a supporting digital

DOI : https://doi.org/10.24952/lavoisier.v5i1.19662

tool in the experimental class to measure and record students' progress in self-directed learning on a regular basis (real-time assessment) throughout the intervention process. All treatments in both the experimental and control classes were implemented over a specific period aligned with the schedule of the Semester Learning Plan (SLP) for the Chemistry English course.

All collected data were statistically analyzed using IBM SPSS Statistics 27 software. The data analysis process began with descriptive statistics to examine the characteristics of the data distribution, followed by tests of analytical assumptions, including the Shapiro–Wilk normality test (given that the sample size per class was < 50 students) and Levene's test of homogeneity. Hypothesis testing was conducted using an independent samples t-test on the mean posttest scores or gain scores at a significance level of 0.05. To measure the effectiveness or strength of the treatment's impact, effect sizes were calculated using Cohen's d formula, classified into three theoretical categories: small (0.20), moderate (0.50), and large (≥ 0.80).

3. Results and Discussions

Results

Descriptive Statistical Analysis of TOEFL Scores

The TOEFL score categories for the Structure and Written Expression section are determined based on the mean score and ideal standard deviation of the total 40 items. These scores are classified into three categories: high (≥ 12.92), moderate (6.40–12.91), and low (< 6.40). A summary of the descriptive statistical analysis results for the control and experimental groups is presented in Table 1.

Table 1. Descriptive Statistics of Pre-test and Post-test TOEFL Scores

Groups	Test Type	N	Mean	Standard Deviation	Category
Control	Pre-test	35	10,286	1,4465	Moderate
	Post-test	35	23,229	1,5546	High
Eksperimental	Pre-test	38	11,800	1,2319	Moderate
	Post-test	38	28,829	1,3170	High

Based on Table 1, the control class achieved an average pre-test score of 10.286 with a standard deviation of 1.4465, which falls into the moderate category. This indicates that the students' initial abilities prior to the intervention were at a comparable level and tended to be homogeneous. After participating in conventional learning, the control class's average post-test score increased to 23.229 with a standard

DOI : https://doi.org/10.24952/lavoisier.v5i1.19662

deviation of 1.5546. This slight increase in the standard deviation indicates that the distribution of post-test scores in the control class became slightly more varied (heterogeneous).

In contrast, the experimental class recorded an average pre-test score of 11.800 with a standard deviation of 1.2319 (moderate category), indicating initial proficiency levels not significantly different from those of the control class. However, after implementing the deep learning approach supported by the EnglishScore platform, the experimental class's average post-test score surged significantly to 28.829. At the same time, the standard deviation decreased to 1.3170. This decrease in the standard deviation demonstrates that the TOEFL scores of students in the experimental class became highly homogeneous and clustered around a high mean.

Assumption Prerequisites Testing

To ensure the validity of parametric inferential statistical analysis, prerequisite tests were conducted, including tests for normality and homogeneity. The normality of the data was assessed using the Shapiro-Wilk test, given that the sample size for each class was less than 50 subjects ($N < 50$). The results of the normality test are summarized in Table 2.

Table 2. Tests of Normality (Shapiro-Wilk)

Group	*Sig. Pretest*	*Sig. Postest*	*Conclusion*
Control	0,138	0,130	Normally Distributed
Eksperimental	0,059	0,052	Normally Distributed

As shown in Table 2, the p-values for the pre-test and post-test data in both the control and experimental classes were consistently above the significance level of $\alpha = 0.05$ ($p > 0.05$). Thus, it can be concluded that all data sets are normally distributed.

Next, the homogeneity of variances between groups was tested using Levene's Test, as presented in Table 3.

Table 3. Test of Homogeneity of Variances (Levene's Test)

	Sig. Pretest	*Sig. Postest*	*Conclusion*
Levene's test	0,277	0,403	Homogeneous

The results in Table 3 show that the p-values for the pre-test data ($p = 0.277$) and post-test data ($p = 0.403$) are greater than 0.05 ($p > 0.05$). This confirms that the variance in the data across groups is equivalent or homogeneous.

Hypothesis Testing

DOI : https://doi.org/10.24952/lavoisier.v5i1.19662

Since all prerequisite assumptions have been fully met, the analysis proceeds to the hypothesis testing stage using an independent samples t-test to assess the significance of the difference in post-test scores between the two groups.

Table 4. Summary of Independent Sample t-test Results

Variable	*t*	*df*	*Sig. (2-Tailed)*	*Mean Difference*	*Cohen's d*
TOEFL Post-test Score	-16.200	71	0,001	5.600	3.86 (Large)

The results of the analysis in Table 4 indicate a highly significant difference between the two groups, with a t-value of 16.542 and a p-value of 0.001 ($p < 0.05$). Thus, the null hypothesis (H0) is rejected, proving that the application of the EnglishScore-assisted deep learning approach has a statistically significant impact compared to conventional methods. Additionally, the effect size test yielded a Cohen's d value of 3.86, which falls into the "large effect" category (≥ 0.80).

Discussions

The Synergistic Impact of Deep Learning and EnglishScore on TOEFL Mastery

The findings of this study provide empirical evidence that the deep learning approach, supported by the EnglishScore platform, optimally improves students' TOEFL scores, particularly in the "Structure and Written Expression" section of the English for Chemistry course. The clear superiority of the experimental class demonstrated by a post-test mean of 28.829 with a highly homogeneous score distribution (SD = 1.3170) compared to the control class (Mean = 23.229; SD = 1.5546) indicates that the integration of digital assessment into the deep learning framework is capable of creating a robust learning ecosystem.

Theoretically, these findings align with contemporary literature on deep learning, which emphasizes that high-level cognitive engagement is essential for fostering long-term retention of understanding (Reimers, 2020). Through a deep learning approach, students no longer merely memorize grammatical patterns or formulas in a mechanistic manner (surface learning). Instead, they are trained to internalize the internal logic of the English language contextually. This supports the argument by (Ridho et al., 2025) who emphasize that exploration- and context-based learning activities can significantly boost students' academic performance in the field of chemistry.

Mastery Cognitive Justification of Science Students Utilizing Mobile Assessment

A deeper analysis of the data reveals why this approach is so effective for chemistry majors. Chemistry students naturally possess a structured, logical, and analytical mindset. When deep learning strategies are combined with EnglishScore, a cognitive synergy is created. The EnglishScore app functions not merely as a static testing tool, but as a platform that provides real-time feedback.

DOI : https://doi.org/10.24952/lavoisier.v5i1.19662

This instant feedback mechanism acts as the primary catalyst for the critical reflection phase in the deep learning process. Chemistry students can immediately identify where their grammatical errors lie in the Structure section, formulate their own hypotheses regarding correct linguistic rules, and immediately test them again through the application. This process confirms the view of Ridho & Ikhsan (2018) that the accuracy of selecting adaptive learning media is linearly correlated with the efficiency of learning outcomes. This continuous cycle of automatic feedback successfully shifts students' orientation from merely "getting a grade" to truly "dissecting the mechanical structure" of the language.

Educational Implications and Limitations

From an academic perspective, this study makes a significant contribution to the literature on Mobile-Assisted Language Learning (MALL) by demonstrating how technology can be optimized to facilitate deep cognitive processing in non-language majors (science). Practically, the success of this model provides a clear blueprint for college instructors to design targeted English for Specific Purposes (ESP) curricula. This framework realizes the essence of the Merdeka Belajar curriculum through differentiated, technology-based learning, which successfully fosters 21st-century competencies among undergraduate chemistry students.

4. Conclusions

Based on the results of the data analysis and discussion, it can be concluded that the implementation of the deep learning approach using the EnglishScore platform has a positive and significant effect on improving students' TOEFL scores in the Structure and Written Expression section. Descriptively, the experimental class achieved a higher average posttest score (28.829) with a more homogeneous score distribution compared to the control class (23.229). The results of the independent samples t-test reinforce this finding with a significance value of 0.001 ($p < 0.05$), indicating a significant difference in achievement between the two groups.

This success is driven by the characteristics of deep learning, which facilitate high-level cognitive engagement and active reflection among students, ensuring that their understanding of English grammatical structures goes beyond mere memorization to a strong conceptual understanding. The integration of the EnglishScore platform has proven effective as an adaptive learning tool that enhances student motivation and independence. Practically, this study has implications for the development of English teaching strategies in science education, suggesting the adoption of digital technology and process-oriented approaches to achieve more optimal and equitable learning outcomes.

References

Bond, M., Bedenlier, S., Marín, V. I., & Händel, M. (2021). Emergency Remote Teaching in Higher Education: Mapping the First Global Online Semester. *International Journal of Educational Technology in Higher Education*, *18*(50). https://doi.org/10.1186/s41239-021-00282-x

Chen, J., & Moradi, H. (2023). Asian English Language Classrooms: Where Theory and Practice Meet. Applied *Linguistics*, *44*(3), 592–595. https://doi.org/10.1093/applin/amaa040

Han, C., & Gardner, S. (2021). However and other transitions in the Han CH-EN corpus. *Journal of English for Academic Purposes*, *51*, 100984. https://doi.org/10.1016/j.jeap.2021.100984

Holmes, W., & Tuomi, I. (2022). State of the Art in AI and the Future of Education. *European Journal of Education*, *57*(4), 542–570. https://doi.org/10.1111/ejed.12533

Hossain, A., & Munam, A. M. (2022). Factors Influencing Facebook Addiction among Varendra University Students in the Lockdown during the COVID-19 Outbreak. *Computers in Human Behavior Reports*, *6*, 100181. https://doi.org/10.1016/j.chbr.2022.100181

Hyland, K. (2022). English for Specific Purposes: What is It and Where is It Taking Us? *ESP Today*, *10*(2), 202–220. https://doi.org/10.18485/esptoday.2022.10.2.1

Kormos, J. (2020). Specific Learning Difficulties in Second Language Learning and Teaching. *Language Teaching*, *53*, 129–143. https://doi.org/10.1017/S0261444819000442

Markauskaite, L., Marrone, R., Poquet, O., Knight, S., Martinez-Maldonado, R., Howard, S., Tondeur, J., De Laat, M., Buckingham Shum, S., Gašević, D., & Siemens, G. (2022). Rethinking the Entwinement between Artificial Intelligence and Human Learning: What Capabilities do Learners Need for a World with AI? *Computers and Education: Artificial Intelligence*, *3*, 100056. https://doi.org/10.1016/j.caeai.2022.100056

Phillips, D. (2001). *Longman complete course for the TOEFL test: Preparation for the computer and paper tests*. Longman.

Reimers, F. M. (2020). *Educating Students to Improve the World*. Springer Singapore. https://doi.org/https://doi.org/10.1007/978-981-15-3887-2

Ridho, D., & Ikhsan, J. (2018). Learning Outcome of Students on Chemondro Media (Chemistry on Android) with Flipped Classroom Model on Electrolyte Solution. *Proceedings of the 3rd Annual International Seminar on Transformative Education and Educational Leadership (AISTEEL 2018)*, 591–595. https://doi.org/10.2991/aisteel-18.2018.129

Ridho, D., Tua, F., Panggabean, M., Febriana Gea, O., Panjaitan, K. A., Daeli, D., & Nadira, F. (2025). Pengaruh Model Pembelajaran Student Centered Learning (SCL) terhadap Miskonsepsi Kimia Unsur Menggunakan Metode Certainty of Response Index (CRI) dengan Model Two-Tier Test pada Mahasiswa Pendidikan Kimia. *Jurnal Ilmiah Wahana Pendidikan, Oktober*, *11*(10.A), 123–133. https://jurnal.peneliti.net/index.php/JIWP/article/view/13038.

Trigwell, K., & Prosser, M. (2020). *Exploring University Teaching and Learning*. Palgrave Pivot. https://doi.org/https://doi.org/10.1007/978-3-030-50830-2

Uzorka, A., & Odebiyi, O. A. (2025). Impact of Digital Learning Tools on Student Engagement and Achievement. *Journal of Digital Learning and Distance Education*, *4*(1), 1436–1445. https://doi.org/10.56778/jdlde.v4i1.511

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1). https://doi.org/10.1186/s41239-019-0171-0

DOI : https://doi.org/10.24952/lavoisier.v5i1.19662