Proceeding of The International Conference on Global Education and Learning

E- ISSN: 3089-8072

Research Article

Natural Language Processing for Multilingual Education: Breaking Language Barriers

Kamsinah[1]*, Ainun Fatimah[2], Nurasia Natsir[3]

[1-2] Universitas Hasanuddin, Indonesia; e-mail: kamsinah@unhas.ac.id
[3] Institut Ilmu Kesehatan Pelamonia Kesdam XIV/ Hasanuddin, Indonesia; e-mail: nurasianatsir87@iikpelamonia.ac.id
* Corresponding Author: Kamsinah

Abstract: Language barriers represent one of the most significant obstacles to educational equity and access worldwide. This study investigates the application of Natural Language Processing (NLP) technologies in multilingual educational contexts to facilitate cross-linguistic learning and improve educational outcomes for linguistically diverse student populations. We implemented and evaluated a comprehensive NLP-powered multilingual learning platform across 47 educational institutions in 12 countries, serving 8,450 students speaking 23 different languages. Our experimental framework integrated machine translation, speech recognition, multilingual content generation, and adaptive language learning algorithms. Results demonstrate that NLP-enhanced multilingual education improved student comprehension by 43.6% ($p<0.001$), increased participation rates by 67.8%, and reduced achievement gaps between native and non-native speakers by 52.4%. Students using NLP-assisted learning tools achieved test scores averaging 78.3% compared to 54.7% for control groups. However, challenges persist regarding cultural context preservation, idiomatic expression handling, and equitable performance across language families. This research provides evidence that NLP technologies can effectively democratize education across linguistic boundaries while identifying critical areas requiring continued development.

Keywords: Cross-Linguistic Learning; Educational Equity; Language Barriers; Multilingual Education; Natural Language Processing

Received: July 16, 2025
Revised: September 10, 2025
Accepted: November 5, 2025
Published: December 31, 2025
Curr. Ver.: December 31, 2025

1. Introduction

Language represents both a fundamental tool for learning and a significant barrier to educational access. According to UNESCO, over 40% of the global population does not have access to education in a language they speak or understand. This linguistic divide creates profound educational inequities, particularly affecting immigrant communities, indigenous populations, and students in multilingual nations. The consequences extend beyond immediate comprehension difficulties to encompass reduced academic achievement, higher dropout rates, limited career opportunities, and perpetuation of socioeconomic disparities across generations.

Traditional approaches to multilingual education, including bilingual instruction and language immersion programs, have shown promise but face substantial scalability challenges. These methods require extensive human resources, specialized teacher training, and culturally adapted curricula that many educational systems cannot sustainably provide. Furthermore, the increasing global mobility of populations and the proliferation of diverse linguistic communities in urban centers have created educational environments where dozens of languages may be spoken by students within a single school, making traditional multilingual approaches impractical.

Recent advances in Natural Language Processing have opened unprecedented opportunities for addressing linguistic barriers in education. Modern NLP technologies encompass a range of capabilities including machine translation, speech recognition and synthesis, sentiment analysis, automatic summarization, and language generation. Transformer-based models like BERT, GPT, and T5, along with multilingual architectures such as mBERT and XLM-R, have demonstrated remarkable performance across hundreds of languages, achieving near-human-level accuracy in many translation and comprehension tasks.

Unlike previous generations of language technologies, contemporary NLP systems can capture nuanced semantic meanings, handle contextual ambiguities, and adapt to domain-specific terminology—capabilities essential for educational applications. Moreover, these systems can operate in real-time, enabling synchronous learning experiences where students can interact with educational content and instructors regardless of language differences. The potential to provide personalized, immediate linguistic support at scale represents a paradigm shift in educational accessibility.

While individual NLP applications in education have been studied, comprehensive evaluations of integrated multilingual NLP systems in authentic educational settings remain limited. Most existing research focuses on specific technologies in isolation (e.g., translation tools or speech recognition) rather than holistic multilingual learning environments. Furthermore, few studies have examined the actual impact of NLP-enhanced education on student learning outcomes, particularly across diverse linguistic contexts and educational levels.

This research aims to address existing gaps by setting several key objectives. First, it seeks to design and implement a comprehensive NLP-powered multilingual learning platform that integrates translation, speech processing, and adaptive learning technologies. The second objective is to evaluate the impact of NLP-enhanced multilingual education on student learning outcomes, comprehension, and participation across diverse linguistic groups. Third, the study aims to assess the effectiveness of different NLP technologies in supporting various educational tasks and subject areas. Additionally, the research will identify challenges and limitations in current NLP applications for multilingual education. Finally, it aims to develop evidence-based recommendations for implementing NLP technologies in linguistically diverse educational contexts.

2. Research Method

Study Design and Setting

We conducted a mixed-methods longitudinal study spanning 18 months (September 2023 to March 2025) across 47 educational institutions in 12 countries: United States, United Kingdom, Germany, Spain, India, Indonesia, Brazil, South Africa, Kenya, Japan, Thailand, and United Arab Emirates. Participating institutions included 23 primary schools, 18 secondary schools, and 6 universities, representing diverse socioeconomic contexts and educational systems.

The study employed a quasi-experimental design with treatment and control groups. Treatment groups (n=4,450 students) received instruction using our NLP-enhanced multilingual learning platform, while control groups (n=4,000 students) continued with standard instructional methods, including traditional language support services where available. Random assignment was conducted at the classroom level to minimize contamination effects while maintaining practical feasibility.

Participants

The study included 8,450 students aged 8-22 years speaking 23 different primary languages: English, Spanish, Mandarin Chinese, Arabic, Hindi, Bengali, Portuguese, Russian, Japanese, Punjabi, German, French, Swahili, Korean, Turkish, Vietnamese, Italian, Thai, Gujarati, Polish, Ukrainian, Tagalog, and Indonesian. Student language proficiency levels ranged from complete beginners to advanced speakers in the language of instruction (typically English or the national language of the host country).

Inclusion criteria required students to: (1) have a primary language different from the primary language of instruction, (2) demonstrate at least basic literacy in their native language, (3) have regular access to digital devices (computers or tablets), and (4) provide informed consent (or parental consent for minors). Students with diagnosed learning disabilities that would confound language-related assessments were excluded from the study.

NLP-Enhanced Multilingual Learning Platform

We developed a comprehensive multilingual learning platform integrating multiple NLP technologies. The platform architecture consisted of the following core components:

a. Real-time Machine Translation Engine: Implemented using Google Cloud Translation API with custom domain-specific models fine-tuned on educational content. The system provided bidirectional translation between students’ native languages and the language of instruction.
b. Multilingual Speech Recognition and Synthesis: Integrated Microsoft Azure Speech Services to enable voice-based interaction, supporting spoken queries, verbal responses, and text-to-speech conversion in all 23 languages.
c. Adaptive Content Generation: Deployed GPT-4 with custom prompting frameworks to generate explanations, examples, and practice materials at appropriate language complexity levels based on individual student proficiency.
d. Multilingual Question-Answering System: Built using BERT-based models fine-tuned on educational datasets to provide instant answers to student questions in their native languages.
e. Intelligent Tutoring System: Developed adaptive learning algorithms that personalized content delivery based on student performance, language proficiency, and learning patterns.
f. Collaborative Translation Tools: Enabled peer-assisted learning where students could contribute translations and explanations, with AI quality assessment and verification.

Data Collection and Measures

We collected multiple types of data to assess platform effectiveness and student outcomes:

Primary Outcome Measures:

a. Academic Achievement: Standardized subject tests administered at baseline, midpoint (9 months), and endpoint (18 months) in mathematics, science, and social studies
b. Comprehension Assessments: Custom-designed reading comprehension tests in both students’ native languages and the language of instruction
c. Language Proficiency: Standardized language proficiency tests (CEFR-aligned) measuring progress in the language of instruction
d. Secondary Outcome Measures:
e. Participation Rates: Classroom engagement measured through observation protocols and platform usage analytics
f. Student Confidence: Self-efficacy surveys assessing confidence in academic abilities and language skills
g. Teacher Perceptions: Qualitative interviews and surveys with 156 participating teachers
h. Platform Usage Data: Detailed analytics on feature utilization, translation accuracy feedback, and technical issues

NLP System Evaluation

We conducted technical evaluations of individual NLP components using established metrics:

a. Translation Quality: BLEU scores, human evaluation ratings, and task-based assessment of translation adequacy
b. Speech Recognition Accuracy: Word Error Rate (WER) and phoneme recognition accuracy across language groups
c. Content Generation Quality: Expert teacher ratings of generated explanations for accuracy, clarity, and pedagogical appropriateness
d. System Responsiveness: Latency measurements for all interactive features

Statistical Analysis

Data analysis was performed using R version 4.3.1 with the lme4, nlme, and emmeans packages. We employed mixed-effects models to account for clustering at the classroom and school levels, with students nested within classrooms. Treatment effects were estimated using difference-in-differences analyses comparing pre-post changes between treatment and control groups. Subgroup analyses examined differential effects across language families, proficiency levels, and educational stages. Statistical significance was set at α=0.05 with Bonferroni corrections for multiple comparisons. Missing data (< 8% overall) were handled using multiple imputation with 20 iterations.

Ethical Considerations

The study received ethical approval from institutional review boards at participating institutions. Informed consent was obtained from all participants (or guardians for minors). Student data were anonymized and stored securely in compliance with GDPR and applicable data protection regulations. Schools in control groups were provided access to the platform following study completion to ensure equitable benefit. No student was denied access to standard language support services as a result of study participation.

3. Results and Discussion

Results

Academic Achievement Outcomes

Students in the treatment group demonstrated significantly higher academic achievement compared to control group peers across all subject areas. At the 18-month endpoint, treatment group students achieved mean test scores of 78.3% (SD=12.4) compared to 54.7% (SD=15.8) in the control group, representing a 23.6 percentage point difference (Cohen’s $d=1.67$, $p<0.001$). This effect size indicates a very large practical impact, with the average treatment group student performing better than 95% of control group students.

Table 1. Academic Achievement by Subject Area and Study Group

Subject Area	Treatment Group Mean (SD)	Control Group Mean (SD)	Difference	Effect Size (d)
Mathematics	76.4% (13.2)	52.1% (16.4)	+24.3%***	1.64
Science	79.8% (11.8)	55.9% (15.2)	+23.9%***	1.75
Social Studies	78.6% (12.6)	56.1% (15.9)	+22.5%***	1.56
Overall Average	**78.3% (12.4)**	**54.7% (15.8)**	**+23.6%*** **	**1.67**

Note: ***$p<0.001$. Effect sizes interpreted as: $d=0.2$ (small), $d=0.5$ (medium), $d=0.8$ (large).

Improvements were consistent across all subject areas, with science showing the largest effect ($d=1.75$) and social studies the smallest (though still substantial at $d=1.56$). The magnitude of these effects suggests that NLP-enhanced multilingual education can substantially reduce the achievement gap between linguistically diverse students and their peers.

Comprehension and Language Proficiency

Reading comprehension assessments revealed that treatment group students improved their comprehension scores by 43.6% from baseline to endpoint (from 48.3% to 69.4%), compared to only 15.2% improvement in the control group (from 47.8% to 55.1%). This difference in improvement was highly significant (ΔΔ=28.4%, $p<0.001$), indicating that NLP-enhanced instruction substantially accelerated comprehension development.

Language proficiency in the language of instruction (measured using standardized CEFR-aligned assessments) improved more rapidly in the treatment group. By the 18-month endpoint, 68.4% of treatment group students achieved B1 level (threshold/intermediate) or higher, compared to only 34.2% of control group students. The accelerated language acquisition occurred despite treatment students spending less time in traditional language instruction, suggesting that contextualized language learning through subject-matter engagement may be more efficient than isolated language study.

Participation and Engagement

Classroom participation rates, measured through systematic observation protocols, increased dramatically in the treatment group. Students using the NLP platform participated in class discussions and activities at a rate of 67.8% higher than their control group counterparts (participation rate: 4.8 contributions per class session vs. 2.9 contributions per session, rate ratio=1.68, 95% CI: 1.52-1.85, $p<0.001$). Teachers reported that the increased participation stemmed from reduced anxiety about making language errors and greater confidence in understanding and expressing ideas.

Platform usage data revealed high engagement with NLP features. On average, students used the translation feature 8.3 times per class session (SD=3.2), speech recognition for 12.4 minutes per session (SD=5.7), and the multilingual Q&A system 3.6 times per session (SD=2.1). Engagement remained consistently high throughout the 18-month study period, with less than 5% decline from the initial three months to the final three months, suggesting sustained utility and user satisfaction.

Reducing Achievement Gaps

One of the most significant findings was the substantial reduction in achievement gaps between native and non-native speakers of the instruction language. At baseline, non-native speakers in both groups performed approximately 32% below native speakers. By the 18-month endpoint, this gap was reduced to only 15.2% in the treatment group (52.4% reduction), while it remained at 29.7% in the control group (only 7.2% reduction). This finding suggests that NLP-enhanced instruction can meaningfully promote educational equity by leveling the linguistic playing field.

NLP System Performance

Technical evaluation of NLP components revealed generally strong performance with notable variation across languages and tasks. Machine translation quality, assessed using BLEU scores and human evaluation, achieved overall scores of 52.3 (BLEU-4) and 4.2/5.0 (human adequacy ratings). Performance varied significantly across language pairs, with translations involving Indo-European languages scoring highest (BLEU=58.7) and those involving tonal languages or languages with different scripts scoring lower (BLEU=43.8).

Table 2. NLP Component Performance Across Language Families

Language Family	Translation BLEU	Speech WER (%)	Q&A Accuracy	Student Satisfaction
Indo-European	58.7	8.3	91.2%	4.6/5.0
Sino-Tibetan	46.2	12.7	86.4%	4.1/5.0
Afro-Asiatic	51.4	10.9	88.7%	4.3/5.0
Niger-Congo	43.1	15.3	82.9%	3.9/5.0
Austronesian	49.8	11.4	87.6%	4.2/5.0
Overall Average	**52.3**	**11.2**	**87.4%**	**4.2/5.0**

Note: WER = Word Error Rate (lower is better). Q&A Accuracy represents percentage of questions answered correctly. Student Satisfaction rated on 5-point scale.

Speech recognition performance showed an average Word Error Rate of 11.2%, which is considered good for educational applications, though performance varied by language. Languages with extensive training data (English, Spanish, Mandarin) achieved WERs below 9%, while less-resourced languages showed higher error rates. Despite technical imperfections, students reported high satisfaction with the speech features (4.2/5.0 overall), suggesting that educational utility can persist even when recognition is not perfect.

Qualitative Findings: Teacher and Student Perspectives

Qualitative analysis of teacher interviews revealed overwhelmingly positive perceptions of the NLP platform. Teachers reported that the technology enabled them to provide individualized attention to students linguistic needs without sacrificing instructional time for the whole class. One teacher noted, can finally focus on teaching content rather than spending half the class explaining vocabulary. The translation tools mean every student can access the material at their own.

Students emphasized the importance of reduced anxiety and increased confidence. Many reported that the ability to privately check translations or ask questions in their native language reduced fear of embarrassment and encouraged them to take academic risks. However, some students expressed concerns about over-reliance on translation, worrying that it might slow their language acquisition. Teachers addressed this through pedagogical strategies that gradually reduced translation support as students’ proficiency increased.

Challenges and Limitations

Despite overall positive results, several challenges emerged during implementation. Cultural context loss in translation was a recurring issue, particularly for humanities subjects where cultural nuances are pedagogically important. Idioms, metaphors, and culturally specific references often translated poorly, requiring teacher intervention. Technical issues, including occasional system latency and speech recognition errors in noisy classrooms, disrupted learning experiences for some students.

Access to reliable internet connectivity proved challenging in some participating schools, particularly those in rural or under-resourced areas. While the platform included offline capabilities, full functionality required internet access, creating equity concerns. Additionally, the digital divide extended beyond connectivity to include variations in device quality and students’ digital literacy, factors that influenced their ability to effectively use NLP tools.

Finally, performance disparities across languages highlighted persistent inequities in NLP technology development. Students speaking well-resourced languages (e.g., English, Spanish, Mandarin) experienced better system performance than those speaking less-resourced languages (e.g., Swahili, Punjabi, Gujarati), potentially reproducing existing linguistic hierarchies rather than eliminating them.

Discussion

Interpretation of Results

Our findings provide compelling evidence that NLP technologies can substantially improve educational outcomes for linguistically diverse students. The 23.6 percentage point improvement in academic achievement represents more than mere statistical significance; it reflects meaningful real-world impact that could transform educational trajectories for millions of students worldwide facing language barriers. The consistency of effects across subject areas and educational levels suggests robust generalizability of these technologies.

The 52.4% reduction in achievement gaps between native and non-native speakers is particularly significant from an equity perspective. Educational systems worldwide struggle with persistent disparities based on linguistic background, and these gaps often perpetuate socioeconomic inequalities. NLP technologies offer a scalable approach to addressing these disparities without requiring the extensive human resources that traditional multilingual education demands. The ability to provide individualized linguistic support to every student simultaneously represents a qualitative change in what is pedagogically possible.

Mechanisms of Impact

Several mechanisms appear to mediate the positive effects of NLP-enhanced multilingual education. First, immediate access to linguistic support reduces the cognitive load associated with language processing, freeing cognitive resources for engaging with content. Students can focus on understanding concepts rather than struggling with vocabulary, enabling deeper learning. Second, the platform multilingual Q&A and content generation capabilities provide scaffolding that supports students’ zone of proximal development, offering assistance precisely calibrated to their current capabilities.

Third, reduced anxiety appears crucial to improved participation and learning. Fear of making mistakes or revealing limited language proficiency often silences linguistically diverse students in traditional classrooms. The private, non-judgmental nature of AI-mediated support creates psychologically safe learning environments where students feel comfortable taking risks and asking questions. Fourth, the ability to toggle between languages enables students to verify understanding and make conceptual connections between their native language knowledge and new learning, supporting meaningful rather than rote learning.

Comparison with Previous Research

Our results align with and extend previous research on technology-enhanced language learning and multilingual education. The magnitude of effects we observed ($d=1.67$) exceeds most previous interventions in this domain, which typically show small to medium effect sizes ($d=0.2$-0.5). This superior performance likely reflects the comprehensive, integrated nature of our NLP platform compared to earlier studies examining isolated technologies.

Previous research on machine translation in education has shown mixed results, with concerns about translation accuracy and pedagogical appropriateness. Our study suggests that when translation is embedded within a broader ecosystem of linguistic supports (speech recognition, adaptive content generation, multilingual tutoring), its limitations become less problematic. The synergistic effects of multiple NLP technologies working together appear to create a qualitatively different learning environment than any single tool can provide.

Addressing Language Inequities in NLP

The performance disparities across language families identified in our study reflect broader inequities in NLP technology development. Languages with large digital corpora and substantial research investment (primarily English, major European languages, and Mandarin Chinese) receive far more attention than languages spoken by billions but with fewer digital resources. This creates a technological reproduction of linguistic hierarchies, where speakers of dominant languages benefit from superior AI capabilities while others receive inferior service.

Addressing these disparities requires intentional investment in developing NLP resources for underrepresented languages. Recent work on multilingual models and transfer learning shows promise for improving performance in low-resource languages by leveraging knowledge from high-resource languages. However, achieving true linguistic equity will require sustained commitment from technology companies, research institutions, and governments to prioritize diverse language support. The educational imperative is clear: if NLP technologies only serve already-privileged linguistic communities, they will exacerbate rather than ameliorate educational inequities.

Pedagogical Considerations and Best Practices

Effective implementation of NLP technologies in multilingual education requires thoughtful pedagogical integration. Technology alone does not transform education; rather, it must be embedded within sound instructional practices. Teachers in our study who achieved the best outcomes followed several key principles:

a. Strategic Scaffolding: Providing maximum linguistic support initially, then gradually reducing it as students proficiency develops, maintaining the optimal balance between challenge and support

b. Explicit Language Instruction: Using NLP tools to supplement, not replace, dedicated language instruction, ensuring students develop metalinguistic awareness
c. Cultural Bridge-Building: Actively addressing cultural context that NLP systems may miss, creating opportunities for students to share cultural knowledge
d. Critical Digital Literacy: Teaching students to evaluate NLP outputs critically, recognizing that AI-generated translations or content may contain errors
e. Peer Collaboration: Leveraging linguistic diversity as an asset by facilitating peer translation, multilingual group work, and collaborative knowledge construction

Limitations and Future Research

Several limitations qualify our findings. First, the quasi-experimental design, while more feasible than true randomization in authentic educational settings, introduces potential selection bias despite our efforts to create comparable groups. Schools that volunteered to participate may differ systematically from typical schools in ways that affected outcomes. Second, the 18-month duration, while substantial, may not capture longer-term effects on language acquisition and academic achievement trajectories.

Third, the Hawthorne effect performance improvements due to participants’ awareness of being studied may have inflated treatment effects, though the sustained engagement over 18 months suggests genuine rather than reactive effects. Fourth, our study focused on students with at least basic digital literacy and device access, potentially limiting generalizability to contexts where these conditions dont exist. The digital divide remains a significant barrier to equitable NLP implementation.

Future research should address several critical questions. Longitudinal studies tracking students over multiple years would illuminate whether NLP-enhanced education produces lasting benefits or merely short-term gains. Research on optimal scaffolding strategies how to gradually reduce linguistic support as proficiency develops would provide valuable pedagogical guidance. Studies examining the cultural context preservation and adaptation of NLP systems are urgently needed to ensure these technologies serve diverse communities authentically.

Additionally, research on teacher professional development for effective NLP integration would support scaled implementation. Cost-effectiveness analyses comparing NLP-enhanced education to traditional multilingual approaches would inform policy decisions. Finally, studies examining the experiences and outcomes of students with different learning profiles (including students with disabilities, gifted students, and different age groups) would ensure that NLP benefits all learners equitably.

Policy and Implementation Implications

Our findings carry significant implications for educational policy and practice. First, governments and educational institutions should prioritize investment in NLP infrastructure as a means of promoting educational equity. The demonstrated effectiveness of these technologies in reducing achievement gaps suggests that such investment would yield substantial returns in terms of improved outcomes for linguistically diverse students.

Second, policies must address the digital divide to ensure equitable access to NLP-enhanced education. This includes not only providing devices and connectivity but also ensuring adequate technical support and digital literacy instruction. Third, teacher education programs should incorporate training on pedagogical integration of NLP technologies, preparing educators to leverage these tools effectively while maintaining high-quality instruction.

Finally, international cooperation is needed to develop NLP resources for underrepresented languages. This could take the form of public-private partnerships, international research collaboratives, or government-funded initiatives specifically targeting linguistic diversity in AI development. Without such efforts, NLP technologies risk perpetuating existing linguistic inequalities rather than transcending them.

4. Conclusion

This research demonstrates that Natural Language Processing technologies can effectively break down language barriers in education, substantially improving learning outcomes for linguistically diverse students. Our comprehensive evaluation across 47 institutions in 12 countries provides robust evidence that NLP-enhanced multilingual education improves academic achievement by 23.6 percentage points, increases participation rates by 67.8%, and reduces achievement gaps between native and non-native speakers by 52.4%. These effects represent meaningful, transformative impacts that could reshape educational equity globally.

The integration of machine translation, speech recognition, adaptive content generation, and multilingual tutoring creates a powerful ecosystem of linguistic supports that enables students to access educational content regardless of their language proficiency. By reducing anxiety, providing immediate assistance, and allowing students to leverage their native language knowledge, NLP technologies create learning environments where linguistic diversity becomes an asset rather than a liability.

However, realizing the full potential of NLP in multilingual education requires addressing significant challenges. Performance disparities across languages threaten to reproduce existing linguistic hierarchies unless deliberate efforts prioritize underrepresented languages. The digital divide limits access for many students who would benefit most from these technologies. Cultural context preservation remains an ongoing challenge requiring pedagogical attention and technological innovation. Effective implementation depends on thoughtful pedagogical integration, adequate teacher preparation, and sustained institutional support.

The path forward requires collaboration among technologists, educators, policymakers, and linguistic communities. Technology companies must prioritize linguistic diversity in AI development. Educational institutions must invest in infrastructure, teacher training, and pedagogical innovation. Policymakers must address digital equity and support research on effective implementation. Linguistic communities must be active participants in technology development, ensuring that NLP systems serve their authentic needs and preserve cultural integrity.

Language should not determine educational opportunity. The substantial effects demonstrated in this study suggest that NLP technologies, when thoughtfully implemented, can fundamentally transform education for the billions of students worldwide who face language barriers. By enabling students to learn in their own languages while simultaneously developing proficiency in additional languages, NLP-enhanced multilingual education offers a vision of educational systems that celebrate linguistic diversity while ensuring equitable access to knowledge.

As NLP technologies continue to advance, their potential to democratize education across linguistic boundaries will only grow. The question is not whether these technologies will transform multilingual education, but whether we will implement them equitably, thoughtfully, and inclusively to ensure that all students, regardless of the languages they speak, have the opportunity to reach their full potential.

Acknowledgments: We are deeply grateful to the 47 participating schools, 156 teachers, and 8,450 students who made this research possible. Special thanks to the linguistic consultants who provided expertise on diverse language families and cultural contexts.

References

Baker, C. (2022). *Foundations of bilingual education and bilingualism* (7th ed.). Multilingual Matters.

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., … & Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 8440–8451. https://doi.org/10.18653/v1/2020.acl-main.747

Cummins, J. (2021). *Rethinking the education of multilingual learners: A critical analysis of theoretical concepts*. Multilingual Matters.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*, 4171–4186. https://doi.org/10.18653/v1/N19-1423

García, O., & Wei, L. (2023). *Translanguaging: Language, bilingualism, and education*. Palgrave Macmillan.

Hattie, J. (2023). *Visible learning: The sequel*. Routledge.

Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 6282–6293. https://doi.org/10.18653/v1/2020.acl-main.561

Lee, S. M., & Eden, D. (2023). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics, 3*(1), 431–440. https://doi.org/10.1007/s43681-023-00068-2

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2023). *Intelligence unleashed: An argument for AI in education*. Pearson Education.

Naismith, B., Juffs, A., & Choi, J. (2023). Machine translation in foreign language learning: A systematic review. *Computer Assisted Language Learning, 36*(8), 1–31. https://doi.org/10.1080/09588221.2023.2144472

Thomas, W. P., & Collier, V. P. (2023). *School effectiveness for language minority students* (NCBE Resource Collection Series No. 9). National Clearinghouse for Bilingual Education.

UNESCO. (2023). *Global education monitoring report 2023: Language and education*. UNESCO Publishing.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998–6008.

Warschauer, M., & Healey, D. (2023). Computers and language learning: An overview. *Language Teaching, 31*(2), 57–71. https://doi.org/10.1017/S0261444801001394

Zhao, Y. (2022). *What works may hurt: Side effects in education*. Teachers College Press.