University of Sydney Journal of TESOL

Volume 4, 2025

ISSN 1834-4712 (Online)

University of Sydney Journal of TESOL

Volume 4, 2025

Aek Phakiti

Editor-in-chief

Marie Stevenson, Phil Chapell and Margaret Kettle

Editors

Published by

The University of Sydney TESOL Research Forum
Sydney School of Education and Social Work
Faculty of Arts and Social Sciences
The University of Sydney

https://journal-of-tesol.sydney.edu.au/

ISSN 1834-4712 (Online)

About the University of Sydney Journal of TESOL

The University of Sydney Journal of TESOL is a peer-reviewed, open-access online journal. Like its forerunner, the *University of Sydney Papers in TESOL*, it is published by the TESOL Research Forum in the Sydney School of Education and Social Work at the University of Sydney, Australia.

This journal offers a platform for postgraduate students, scholars, and practitioners to share their insights, research, and practical experiences in TESOL and Related areas. We aim to foster a community in which the voices of postgraduate students, scholars, teacher trainers, teachers, and other practitioners can be heard.

The journal accepts for submission the following types of contributions:

- reports of original research
- discussions
- pedagogical practices
- interviews with experts
- reviews.

The journal welcomes contributions from a wide range of TESOL and TESOL-related areas. To give some examples:

- bilingualism/bilingual education
- discourse/pragmatics
- English for specific purposes
- intercultural communication
- language testing and assessment
- language policy and planning

- online teaching and learning
- second-language acquisition
- teacher professional development.

SUBMISSIONS OF MANUSCRIPTS

After ensuring they have met the Guidelines for Contributors, authors should submit potential contributions electronically to usydtesoljournal@gmail.com.

Contents

Editorial

Aek Phakiti[1], Marie Stevenson, Phil Chappell and Margaret Kettle

Volume 4 of the *University of Sydney Journal of TESOL* brings together a diverse and thought-provoking collection of contributions that reflect the evolving landscape of English language teaching, learning, and assessment. Across empirical studies, reflective narratives, and conceptual analyses, this volume explores key themes including identity, innovation, interdisciplinarity, and the impact of emerging technologies on Teaching English to Speakers of Other Languages (TESOL) practice.

EXPLORING IDENTITY, INNOVATION, AND INTERDISCIPLINARITY IN TESOL

Michael Burri's article "*Using a Transdisciplinary Lens to Make Sense of My Own Acquisition and Retrieval of Spoken English*" offers a richly layered self-study that draws on neuroscience, psychology, health, and education to explore the phenomenon of language retrieval. Burri's journey from early bilingualism in the USA to language attrition in Switzerland and rapid reacquisition in New Zealand serves as a springboard for pedagogical insights into brain plasticity, emotional safety, and the role of early exposure. His transdisciplinary approach challenges conventional boundaries in TESOL research, inviting educators to consider how scientific understandings of the mind and brain can inform their teaching and learning practices.

[1] Correspondence: Aek Phakiti: aek.phakiti@sydney.edu.au

ISSN 1834-4712 (Online).

DOI: 10.5281/zenodo.17587901

Citation: Phakiti, A., Stevenson, M., Chappell, P., & Kettle, M. (2025). Editorial. *University of Sydney Journal of TESOL*, *4*, 1−4. https://doi.org/10.5281/zenodo.17587901

Diem Tran Thi Ngoc's case study "*Vietnamese EFL University Students' Perceptions and Experiences of Using the ELSA Speak Application for Developing English Oral Fluency*" provides a timely investigation into AI-assisted language learning. Through a combination of screen recordings, learner diaries, and interviews, Diem T. captures the nuanced experiences of Vietnamese university students as they engage with the ELSA Speak app. Her findings highlight the importance of learner autonomy, personalised feedback, and authentic interaction in mobile-assisted language learning (MALL). At the same time, the study raises critical questions about the limitations of AI tools, including issues of feedback accuracy, technical constraints, and pedagogical alignment.

Jack C. Richards' article "*The Hidden Life of the Classroom*" returns us to the heart of language education: the classroom. Through a conceptual exploration of the teacher self and the learner self, Richards illuminates the complex interplay of identity, beliefs, motivation, and agency in shaping classroom dynamics. Drawing on decades of scholarship and practitioner narratives, he argues for reflective practices that uncover the often unseen dimensions of teaching and learning.

NAVIGATING COMPLEXITY IN ASSESSMENT, IDENTITY, AND INNOVATION IN TESOL

Phuong Tran and Cara Dinneen's article "*Reimagining EAP in the Age of GenAI: Innovation, Integrity, and the Future of Assessment*" offers a timely and practice-informed analysis of how English for Academic Purposes (EAP) programs are adapting to the rise of generative artificial intelligence. Drawing on case studies from Macquarie University College and Monash College, the authors illustrate how dual-track assessment regimes combining secure, invigilated checkpoints with guided, AI-assisted learning tasks can preserve academic integrity while fostering digital and AI literacy. Their work highlights the importance of programmatic assessment, genre pedagogy, and feedback literacy in preparing students for the realities of AI-augmented higher education.

Talia Isaacs' article "*The ABCs of Psychometric Validity: A Practical Guide for English Language Teachers"* demystifies the concept of validity in language assessment and makes psychometric principles accessible to classroom teachers. Through clear explanations and practical guides, Isaacs bridges the gap between technical assessment theory and everyday teaching practice. Her discussion of construct, content, criterion-related, and consequential validity—alongside the role of

washback and open science—empowers educators to critically evaluate both their own assessments and external testing regimes.

Mark Fraser's reflective narrative, *"Making the Connections: A Teacher Educator's Perspective on the Compatibility of English Language Teaching Methods,"* offers a deeply personal and pedagogically rich account of teaching TESOL methods to postgraduate students. Fraser foregrounds the role of professional identity and critical reflection in helping students contextualise and adapt teaching methods to their own environments. His development of a critical reflection framework, informed by symbolic interactionism and self-study methodology, provides a powerful tool for fostering meaningful connections between theory and practice.

REFLECTING FORWARD: DIALOGUE, DIVERSITY, AND THE FUTURE OF LANGUAGE EDUCATION

Volume 4 concludes with a thought-provoking multimedia contribution by the Editor-in-Chief, Aek Phakiti. "*Conversation with Professor Constant Leung"* is an eight-part video dialogue that provides a comprehensive overview of language education through the lens of one of the field's most influential scholars. Professor Leung's reflections traverse decades of theoretical development, pedagogical shifts, and assessment reform, culminating in a timely discussion of artificial intelligence and its implications for language learning and evaluation. Leung challenges conventional notions of language proficiency, advocating for a more context-sensitive, socially situated understanding of communicative competence. His critique of dominant models of English and standardised assessment frameworks resonates with current debates on translanguaging, plurilingualism, and inclusive pedagogy. The conversation also explores the ethical and epistemological dimensions of AI in education, advocating for transparency, fairness, and critical engagement with emerging technologies.

CLOSING REMARKS

Taken together, the articles in Volume 4 present a rich tapestry of insights into the contemporary challenges and possibilities of TESOL. From transdisciplinary inquiry and AI-enhanced fluency development to assessment literacy, teacher identity, and critical reflection, this volume affirms the importance of thoughtful, context-aware, and ethically grounded approaches to language education.

We thank all the contributors for their scholarship, creativity, and commitment to advancing the field. We hope this volume inspires further research, reflection, and innovation in TESOL and applied linguistics.

Using a Transdisciplinary Lens to Make Sense of My Own Acquisition and Retrieval of Spoken English

Michael Burri[1]

Wenzhou-Kean University, China, and University of Wollongong, Australia

ABSTRACT

Narratives and personal accounts of language learning experiences provide valuable insights into language development. However, the process by which multilinguals access and retrieve spoken language is not well understood. This paper aims to use a transdisciplinary perspective that draws on the interconnected areas of Mind (psychology), Brain (neuroscience), Health (well-being), and Education to explore my own English language acquisition and retrieval process. I acquired spoken English as a young child in the USA, lost the language while living in Switzerland for 18 years, and then rapidly retrieved it as a 22-year-old English language learner in New Zealand. Taking this transdisciplinary perspective and answering four key questions sheds light on my rather unusual 'spoken English journey.' The paper concludes with a discussion about pedagogical implications for language teachers.

Keywords: bilingualism; language acquisition; language retrieval; spoken English; transdisciplinary

[1] Correspondence: Michael Burri: mburri@kean.edu; https://orcid.org/0000-0003-0115-5349

ISSN 1834-4712 (Online).

DOI: 10.5281/zenodo.17588003

Citation: Burri, M. (2025). Using a transdisciplinary lens to make sense of my own acquisition and retrieval of spoken English. *University of Sydney Journal of TESOL*, *4*, 5–32. https://doi.org/10.5281/zenodo.17588003

INTRODUCTION

English is the most widely spoken language in the world, with roughly 1.5 billion speakers using it in one form or another (Patel et al., 2022). Because the development of English competence has been linked to economic gains and employment benefits, governments worldwide have made English a mandatory subject at primary school level and beyond (e.g., Garton & Copland, 2022; Zein & Butler, 2024). Yet, the exact age at which humans learn an additional language most effectively is an area that continues to be debated.

As a young child living in the USA, I spoke English fluently but was also able to communicate with my mom and dad in Swiss-German. I was a bilingual kid who used two languages regularly (Wodniecka et al., 2020). Right before I turned four, we moved to Switzerland, and as soon as we arrived, my spoken English began to deteriorate. Having to take mandatory English classes in middle school for three years made no difference. By the age of 22, I could no longer retrieve – or easily access – English, and thus was unable to converse in the language I had spoken fluently as a boy. Being in possession of American citizenship, I felt reasonably embarrassed and therefore decided to travel to New Zealand (NZ) with the goal of not only skiing the beautiful Southern Alps, but more importantly, regaining English proficiency. Upon commencing my studies, I was placed in the lowest level at a private language school in Christchurch. However, I soon began to excel and gained relatively high English fluency, with a near-native accent, within a few months.

Why was I able to learn – or more accurately, *retrieve* – spoken English within a few months and attain native-like proficiency in New Zealand at the age of 22? This is a legitimate question given that how bilinguals access and process language is still not well understood (Wodniecka et al., 2020), as well as the proposition that conversational proficiency in an additional language can take two or three years to develop (Cummins, 1999). I have presented at international conferences (e.g., Burri, 2023, July) and written about some of my language learning and professional experiences (Burri, 2022). However, as a teacher educator and researcher, I have always wanted to take a deep dive into neuroscience to help me understand the reasons for initially acquiring spoken English as a young child in the USA, forgetting it in Switzerland, and then retrieving the once forgotten language quite rapidly as an adult in NZ. The opportunity to explore this process arose during my sabbatical in 2024, as a student in the (online) *Neuroscience of Learning* course offered through the Harvard Extension School.

I should point out that the goal of my pursuit of finding answers and writing this paper was not a gratification of my successes as a second language (L2) learner, but rather a desire to contribute to the understanding of language learning and subsequently propose some implications for language teaching based on my own unique experiences.

The need for a transdisciplinary perspective on language acquisition

Narratives of L2 learning experiences (e.g., Hiver et al., 2019) and personal accounts, such as Casanave's (2012) documentation of her informal learning of Japanese, have provided valuable insights into L2 development. However, this L2 learning process differs for individuals who are brought up bilingual from birth, as their brains are slightly different from those of someone acquiring an additional language later in life (Li et al., 2024). Moreover, learning a language in the first few years, only to lose it and then eventually retrieve it, is an under-researched area that warrants further exploration.

Applied linguists have pointed out the importance of transdisciplinarity (e.g., Filipović, 2019; Perrin & Kramsch, 2019) with research combining bilingualism, L2 learning, neuroscience, and applied linguistics (e.g., Cohen, 1980; Fabbro, 1999; Morgan-Short & van Hell, 2023; Nouri, 2015; Suzuki et al., 2023; van Hell, 2023) to understand the process of acquiring one, two, or multiple languages. Yet, few seem to have taken a transdisciplinary perspective that draws on Mind (psychology), Brain (neuroscience), Health (well-being), and Education (MBHE; Tokuhama-Espinosa et al., 2023) to explore one's own language learning endeavours. As such, there is relatively little research on the process of acquiring a language as an adult. Using a transdisciplinary (i.e., MBHE) lens is, therefore, important because it allows me to explore this issue from different but intertwined perspectives and subsequently gather articles, evidence, and answers from multiple fields to help me make sense of my rather unusual 'spoken English journey.' This is, of course, a subjective undertaking based on my own experiences; therefore, if this paper were written by other bilinguals, it would likely look quite different. Nonetheless, I believe a self-study (rather than a typical empirical paper) in which the author takes a transdisciplinary perspective makes a valuable contribution because it may provide important and unique insights into language learning, and, more specifically, into how bilinguals access and process language.

Searching the literature for answers

My attempt to find answers in the *Neuroscience of Learning* course began with identifying key journal articles, book chapters, and books published in high-quality journals and authored by leading scholars within the last decade. Ten years was not necessarily a fixed timeframe, but it allowed me to narrow the search for publications dealing with bilingualism and language acquisition. I made extensive use of Google Scholar and relevant keywords in my search for readings. However, the search was also guided by weekly 'bundles' provided in the online platform of the *Neuroscience of Learning* course. A bundle is a Word document that lists dozens of MBHE resources (e.g., articles, links to YouTube clips) relevant to the weekly topic covered in the course. I first downloaded the bundle and then scanned the titles of all the provided resources. If a reading appeared to be relevant to my topic of investigation, I downloaded it and read the abstract and conclusion. Once done, I either discarded the paper or read it in its entirety, taking notes on key points applicable to my journey of acquiring and improving spoken English. This type of selective, yet in-depth reading of MBHE (i.e., transdisciplinary) literature gradually augmented my understanding of issues surrounding the acquisition and retrieval process of my spoken English. It should be noted, however, that the selected publications were not equally balanced among the four MBHE areas. The subjective nature of my literature search, time constraints, and/or specific use of keywords likely resulted in most publications drawing from Brain (neuroscience) and Education (applied linguistics/language), with fewer articles from Mind (psychology) and Health. While this imbalance was not ideal, the selected publications provided me with a new understanding of how I initially acquired English, lost it, and eventually retrieved it as an adult.

Accessibility of this paper and four key questions

While answers to my question of acquiring and retrieving spoken language were available in the literature, many of the journal articles I read were written in complex, scientific language. Therefore, this paper connects the literature with my own experiences of retrieving spoken English in an attempt to make neuroscience more accessible and meaningful to language teachers and researchers. At the same time, linking my story to MBHE Science may also help refute some specific myths – or false beliefs – about, for example, bilingualism or multilingualism commonly held around the world (Betts et al., 2019; Tan & Amiel, 2022; Tokuhama-Espinosa, 2018).

One broad question that has been on my mind and thus guides this paper is: *How and to what extent did exposure to spoken English in my early years influence the successful retrieval of spoken English after almost two decades of non-use?* Numerous influential, yet somewhat dated, theories offer answers, including Krashen's (1982) Second Language Acquisition Theory, which posits five hypotheses (acquisition-learning, monitor, input, affective filter, and natural order), Long's (1981) Interaction Hypothesis, and Sociocultural Theory (Lantolf & Thorne, 2006). In other words, there are dozens of possible reasons for my English retrieval in NZ. However, based on an initial review of the MBHE literature, I narrowed in on four key questions that I thought related exceptionally well to my own lived experiences, and, ultimately, helped me answer my broad question.

I will now present the four key questions and connect each one to current MBHE research and literature, signifying a transdisciplinary perspective to make sense of my process of acquiring and retrieving spoken English. The key questions include: (1) *Does the Relationship of Brain Plasticity with 'Use it or Lose it' Explain the Ability to Retrieve Spoken Language?;* (2) *Are People Born with an Innate Ability to Retrieve Spoken Language?*; (3) *Does a Child's Language Environment Impact the Retrieval of Spoken Language Later in Life?*; and (4) *Does the Impact of Emotions on Memory Facilitate the Retrieval of Spoken Language?* In each question, key concepts are first defined and discussed, followed by evidence and examples from my own life.

Key question 1: Does the relationship of brain plasticity with 'use it or lose it' explain the ability to retrieve spoken language?

Learning alters the structure of the brain, a phenomenon known as *plasticity* (Costandi, 2016). As the green part in Figure 1 shows, the small gaps at the end of a neuron (brain cell) are called *synapses*; that is, the place where neurons connect with each other through chemical and electrical signals. At these gaps, electric signals pass from one neuron to the next one and neurotransmitters (chemicals) are released. Through practice and repetition, the connections between the neurons are strengthened. Thus, as proposed by Hebb (1949) more than 70 years ago, 'neurons that fire together, wire together' is the physiological mechanism (plasticity) believed to result in learning.

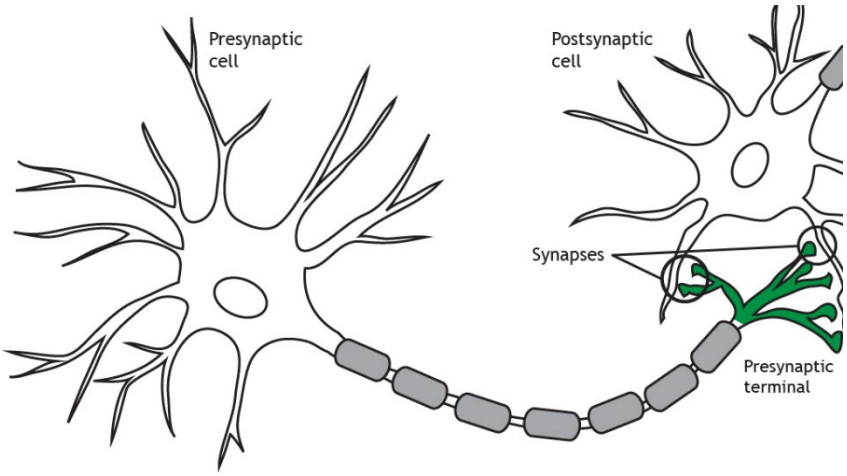


Figure 1. Neurons connecting at synapses

Note: The image was taken from Creative Commons (i.e., open access).

Importantly, while the brain is remarkably plastic during the early years, it continues to change throughout life (Cozolino, 2013; Power & Schlaggar, 2017). That is, plasticity continues "throughout life, in response to everything we do and every experience we have" (Costandi, 2016, p. 2) and "it is likely that millions of synapses are modified in the human brain every second in one way or another" (p. 66). This synaptic modification creates and strengthens neural pathways in the brain, enabling us to improve at what we do, including language learning (Kail & Isel, 2021). As such, the brain adapts to what it does most, but learning something new at an older age is more difficult because we must override old and existing neural pathways with new ones.

Everything works within neural networks – "or basic brain circuitry" (Tokuhama-Espinosa, 2019, p. 1) in the brain. Notably, research has shown that bilingualism shapes the brain's structure (Pliatsikas et al., 2020; van Hell, 2023). Having access to and using two languages appears to increase grey matter density (the number of neurons) and white matter integrity (myelinated axons; see the following paragraph). Changes in brain structure can take place in a relatively short period of time and at all ages, but the process is sensitive to several factors, including similarity of the two languages, individual speaker differences, acquisition age, and gender (Li et al., 2014). However, as Schwieter and Rastelli (2023) point out, "[e]arly bilingualism – a linguistic situation in which an individual is exposed to more than one language from birth – changes a child's brain organisation" (p. 366).

My own experiences with acquiring, forgetting, and retrieving spoken English support Schwieter and Rastelli's suggestion that "*some of these changes persist over the speaker's lifespan*" (p. 366, emphasis added). In combination with bilingual adults' strong ability to control their own behaviour and actions – i.e., their executive control (Bialystock, 2009) – plasticity appears to play an important role in language use and retrieval.

A closer examination of myelination, pruning, and remyelination offers further insights. As the grey parts in Figure 1 show, myelination is a fatty and multi-layered sheath that surrounds the axon of a neuron (Kaller et al., 2017; Tomassy et al., 2016). Myelination not only facilitates the electric signal to travel down the axon faster, thus strengthening the connection and communication between neurons, but it is also caused by repeated practice (rehearsal) of a particular skill or behaviour. Put differently, if the myelin sheath has been sufficiently developed, access and retrieval of stored information in the brain are achieved more quickly (Tokuhama-Espinosa, 2019). It is, therefore, reasonable to assume that being able to speak English fluently in the USA resulted in an increase in white matter (i.e., strong myelination) and long-term competence (Tu et al., 2015), which then allowed me to retrieve my spoken English in NZ.

Pruning occurs when the brain eliminates synapses, and myelin loss occurs because a skill or behaviour is not rehearsed enough or is no longer used (Xin & Chan, 2020). Pruning, therefore, leads to more efficient neural networks in the brain as less real estate is required to complete a task. This is a natural process, and the brain almost always overconnects before it refines its networks. Nevertheless, in terms of language control and use, if a speaker is removed from an enriched language environment and exposure to that language begins to decrease, such as my English in Switzerland, spontaneous production also decreases (Tu et al., 2015). From a plasticity point of view, because I ceased to speak English in Switzerland, the notion of 'use it or lose it' explains why I began to forget English. However, why did my spoken English return so quickly when I moved to New Zealand? Recent research suggests that the synaptic connections remained, but the axons were remyelinated (e.g., Bloom et al., 2022; Neely et al., 2022; Quan et al., 2022), whilst being motivated to learn and spending time in an English immersion environment in NZ. Living in Switzerland meant that I did not have easy access to the language. However, my adult English immersion in NZ reinforced past learning and most likely augmented the remyelination of axons, possibly allowing me to retrieve my spoken English rapidly.

Key question 2: Are people born with an innate ability to retrieve spoken language?

Over the last 100 years, a plethora of theories have been proposed for how first and additional languages are acquired (see Broad, 2020, for a brief overview). The theory most relevant to my second key question is Chomsky's (1965, 2015) hypothesis that all children are born with an innate set of grammatical principles common to all languages (see also Farid et al., 2021). Based on this premise, all humans possess an innate language sense; therefore, certain principles of language do not need to be learned, as they are hardwired into a child's brain and genes.

While Chomsky's proposition has been highly influential, his views have been criticised for ignoring the need to consider the triggers of language development, the social nature of language use, and the individual nature of language development (Cook, 2016). Nevertheless, contemporary neuroscientific understanding supports Chomsky's idea in that genes do "play a substantial role in shaping human behaviour" (Barlow, 2019, p. 74), and neural networks "are inherited through our genes and [then] strengthened through our daily experiences" (Tokuhama-Espinosa, 2019, p. 1). Clearly, genes appear to play a significant role in learning; however, the study of genes is a complex area, and their exact relationship to learning remains poorly understood (Thomas et al., 2015). Yet, because genes are heritable, it could mean that I was born with a specific genetic language makeup and abilities (Eising et al., 2022) that I inherited from my mom and dad. Multilingualism has been a part of my family for generations. My mom, for example, learned English in the USA, but was already fluent in three other languages: Swiss-German, German, and French. It could be that this particular epigenetic tag (Ortuño-Sahagún et al., 2019) – or from a Chomskyan perspective, genetic predisposition to have a gift for languages – enabled me to retrieve my spoken English quite rapidly as an adult in NZ. I may have also inherited a 'good language ear' and/or auditory memory. A heightened sensitivity to different sounds (e.g., the phonemes 'm', 'e', and 't' in *met*) and thus the ability to process auditory input may have not only facilitated my English language learning process (e.g., Saito et al., 2021, 2022) but it might have facilitated language retrieval in NZ.

However, one's genetic make-up does not necessarily determine how well someone acquires a language or accelerates at something. Only 10% of our genes get expressed – or turned on (Cantor, 2021) – because genes, environmental conditions, relationships, and experiences all interact together in shaping observable characteristics (e.g., eye colour, height) and in forming the foundation of brain architecture during the early years of a child (Fox et al, 2010). Thus, I now

turn to the third key question, which focuses on the language environments in the USA and Switzerland.

Key question 3: Does a child's language environment impact the retrieval of spoken language later in life?

The home and language environment in which a child grows up creates specific neural pathways and thus contributes to speech development (Bergelson et al., 2023). Yet, as Thomas and colleagues (2015) discuss, individual differences in motivation, achievement, and abilities result from the environment interacting with an individual's unique genetic profile; therefore, "the same genes may be [shaped] differently in different environments" (p. 75). Hence, this section discusses two specific periods in my early childhood life that may provide insights into the retrieval of my spoken English in NZ: (1) the English and Swiss-German environment in the USA, and (2) the Swiss-German environment in Switzerland.

English and Swiss-German environment in the USA

Evidence from neuroscience suggests that the sensitivity to different sounds of a language is established sometime between five months of gestation and three years of age. Kuhl and colleagues' (2006) research explored the perception of 6- to 12-month-old infants in the USA and Japan, showing that "exposure to a specific language causes neural commitment to the properties of native-language sounds" (p. F19). The researchers suggested that newborns are universal receivers of all language sounds and thus can hear any sounds of any language spoken in the world. However, the study also showed that this extraordinary ability begins to disappear around six months, as the brain starts to focus on the language of the environment in which the baby lives. In other words, the ability to perceive (or hear) sounds and then, ultimately, reproduce them in that particular language begins to be established after about six months of age. Although the notion of a critical period in language learning has been debated (see van Hell, 2023), Kuhl and colleagues' research demonstrated the existence of a specific time period in which both the age of the child and the environment play important roles in acquiring the sound system of a language. As such, learning two or more languages at a younger age when the brain is especially receptive and responsive to specific environmental experiences and exposures shapes bilingual brain architecture (Wei et al., 2015) and leads to changes in brain areas that are responsible for language control (see Tu et al., 2015, for a discussion of these areas).

It is, therefore, reasonable to suggest that being immersed in a social environment in which two languages (English and Swiss-German) were used regularly between my parents, brother, and I – and possibly even hearing the two languages all the time even if no one spoke directly to me in English (i.e., passive exposure; Kurkela et al., 2019) – contributed substantially to my bilingual development (Hoff & Core, 2015). That is, the neural pathways associated with auditory processing (Willmore & King, 2023) were well-rehearsed (i.e., myelinated), and due to substantial exposure to both languages, the neural networks of both my English and Swiss-German auditory systems were established during my four years in the USA. Also, with English and German belonging to the same Indo-European language family and, therefore, by being immersed in two historically similar languages with the same subject-verb-object structure (Tokuhama-Espinosa, 2008) and similar grammar, vocabulary, and sound patterns, my brain may have never pruned neural connections of either language in early childhood. Importantly, "[t]he earlier a second language is learned, the more similar [are] the neural language networks for the two languages" (Friederici, 2017, p. 155), and so memories of two historically similar languages, such as English and Swiss-German, were likely clustered together in neighbouring neural networks (Tokuhama-Espinosa, 2019). Thus, early exposure to both English and Swiss-German seems to have shaped my bilingual brain architecture, which contributed to the quick retrieval of my spoken English in NZ, even though I had not used the language for almost two decades.

Swiss-German environment in Switzerland

As discussed in the introduction, when I arrived in Switzerland, my spoken English began to deteriorate. Due to reduced exposure to English, the activation threshold in my bilingual mind likely began to increase, making access to and control of English more difficult (Tu et al., 2015; Wodniecka et al., 2020). At the same time, the need for group association and language socialisation (Duff & May, 2017) – using the local language to fit in, making Swiss-German-speaking friends, and becoming a member of the Swiss community for the following 18 years – made me favour Swiss-German over English, which, in turn, reduced lexical access and caused language attrition.

Casado and colleagues (2022) suggest that "[i]n speakers who know more than one language, all their languages remain active in their mind even when only one language is required in a given context" (n.p.). The bilingual brain, therefore, always has access to both languages, but it opts for the better choice when a word in two languages needs to be chosen (Sudarshan & Baum, 2019). This seems to be

a particular issue for bilinguals whose languages are not equally balanced because they not only require more effort to shuttle between languages than balanced bilinguals, but they need to work harder to inhibit the more dominant language in order to access and use words in the weaker language (Persici et al., 2019). As far as I remember, in the first few years of being in Switzerland, I was more proficient in English than in Swiss-German. Thus, my motivation to belong in school and, more broadly, the Swiss community might have inhibited my English for me to access, control, and use Swiss-German, resulting in eventually limited or even blocked lexical access in English.

At the same time, reduced exposure to English may have also caused language attrition, a gradual decline in English proficiency (see Schmid & Köpke, 2019). Chaouch-Orozco and Martín-Villena (2025) suggest that reduced exposure to a first language (L1) results in lexical attrition. That is, a bilingual may begin to experience difficulties with accessing and using their L1 words when immersed in an L2 environment. This might mean that being immersed in a Swiss-German environment impacted and restricted my English lexical memory network, thereby limiting my spoken English. My ability to organise and use Swiss-German vocabulary began to improve while my access to English words weakened. This, however, is speculative because numerous words (i.e., cognates) between Swiss-German and English overlap phonologically (e.g., *Bruder* vs *brother*), and also because of Oppenheim and colleagues' (2020) study showing that American children transitioning from Spanish-speaking homes to English-speaking primary schools did not necessarily experience Spanish attrition due to increased English use. A lot, of course, depends on the quantity and quality of language exposure, but in their research, the children's Spanish and English developed concurrently, with more notable increases in English proficiency.

In short, Swiss-German immersion may have made it challenging for me to access and use English in Switzerland. But, having been immersed in a bilingual English–Swiss–German environment in the USA as a young child, these early experiences might have allowed me to reactivate underused English language networks rather than having to create new neural pathways, resulting in relatively quick spoken English retrieval as an adult learner in NZ. Nevertheless, emotions and memory may have also played an important role, and therefore, the fourth key question explores whether the impact of emotions on memory facilitated the retrieval of my spoken English in NZ.

Key question 4: Does the impact of emotions on memory facilitate the retrieval of spoken language?

According to Betts and colleagues (2019), "[e]motions can affect human cognitive processes, including attention, learning and memory, reasoning, and problem-solving" (p. 8). In fact, there is no cognition without emotion (Immordino-Yang & Damasio, 2007). Therefore, emotions are always involved in memory formation, storage, and retrieval (Murray et al., 2013). That is, the physiology of the senses passes through our brain, following basically the same process: sensory perception (e.g., through smell or sight) – emotion – memory – learning. Figure 2 provides a simplified representation of the brain areas involved in emotions and cognitive processes, including the hippocampus, amygdala, and prefrontal cortex. In a nutshell, sensory input – for example, the smell of a fried chicken – travels up the brain stem to the (1) amygdala. The "emotional switch station" (Willis & Willis, 2020, p. 34), located near the centre of our brain, determines whether the smell is something to be fearful of. From there, it moves to (2) the prefrontal cortex, where decisions are made, and then, within a fraction of a second, the signal heads to (3) the hippocampus to confirm with long-term memory whether the smell is indeed fried chicken (Tyng et al., 2017).

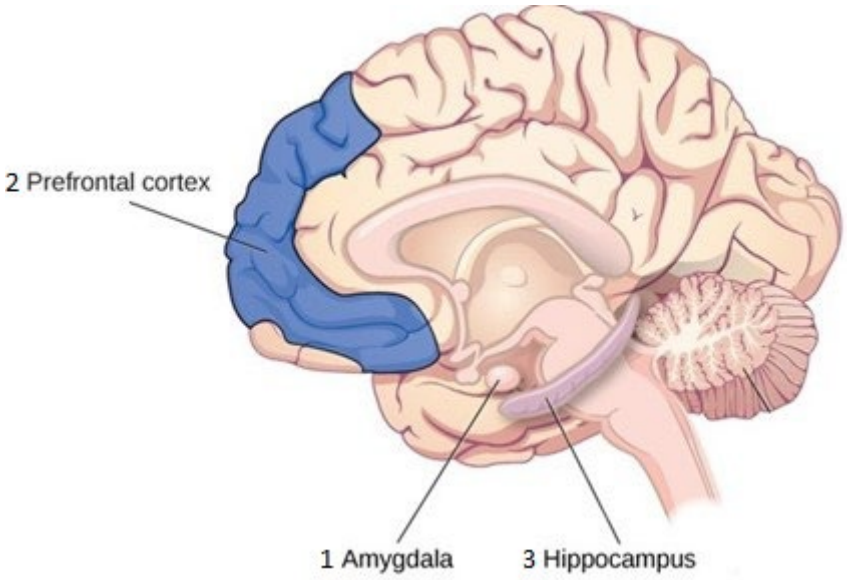


Figure 2. Brain areas involved in emotions and cognitive processes

Note: The image was taken from Creative Commons (i.e., open access).

Emotions are not easily defined (Tyng et al., 2017), and a discussion between the differences of emotions and feelings is beyond the scope of this paper, but if memory is attached to an emotion, the (long-term) storage and retrieval of that memory is enhanced (Dolcos et al., 2020). Emotions are also part of our body, with recent research suggesting that they are controlled by different neurotransmitters (e.g., serotonin, dopamine, endorphins) in the brain (Wang et al., 2020). In other words, emotions are embodied. Embodiment theory suggests that emotions, thought, and language are intricately intertwined with motor processing (Damasio, 2021) – or, in other words, with movement – and thus influence a range of cognitive processes, including learning and meaning-making (Bergen, 2012).

Because language is embodied (Zappa et al., 2024), it plays a crucial role in the formation, storage, and retrieval of emotions and memories. That is, language helps us make sense of sensory input, perception, and experiences. By hearing the word "happiness," for example, language brings together sensory input, emotions, and memory, thus facilitating meaning-making (Lindquist et al., 2015). Both negative and positive emotions impact learning and memory formation (Shao et al., 2020).

Although emotions are embedded faster when they are negative (Salsano et al., 2024), positive emotions resulting from an emotionally safe, social, and linguistically rich English and Swiss-German home environment likely contributed to my bilingual brain development in the USA (Hartas, 2011; Wessels & Trainin, 2023). As discussed in the previous key question, this led to a strengthening in language pathways and as Costandi (2016) postulates, this "synaptic strengthening is indeed necessary and sufficient for memory formation" (p. 60). However, the notion of memory storage at the synaptic level has been questioned recently. Gersham (2023), for instance, argued that memory may be stored in molecules and is altered every time information is recalled, which could mean that language use influences memory formation. Irrespective of where and how memory is stored, being immersed in a safe social-emotional environment (Immordino-Yang et al., 2019) likely contributed to the formation of long-term memories associated with the English language, eventually facilitating the retrieval of spoken English in my early twenties in NZ.

In contrast, stress – a negative emotion – can be detrimental to learning (Shao et al., 2020). Ongoing (chronic) stress increases cortisol (i.e., a stress hormone), which, in turn, interferes with attention and negatively impacts memory. Put simply, stress impedes memory (Trammel & Clore, 2014) and thus learning. I

remember regularly experiencing stress, anxiety, and boredom during my time in middle school in Switzerland, and I did not enjoy the mandatory German, French, and English classes (see Pawlak et al., 2024, for an excellent volume on boredom in the language classroom). Moreover, my motivation to study English was extremely low because all my friends spoke Swiss-German, and with school success being largely measured in Swiss-German, I saw little purpose in studying foreign languages. Attention and memory are the cornerstone of learning (Tokuhama-Espinosa, 2019) and the combination of a lack of purposeful attention and low engagement most likely inhibited my spoken English, so I could use Swiss-German. This led me to 'forget' English until I decided to fly to NZ and enrol in an English language course in Christchurch.

SUMMARY

My four key questions have revealed that plasticity, as well as immersion in two historically similar languages and in an emotionally nurturing and linguistically rich environment, must have played important roles in English acquisition during my early years in the USA and in the eventual retrieval of the language in NZ. It is also possible that I had inherited a certain genetic language ability and/or auditory memory from my parents that then contributed to my ability to retrieve English in NZ. Furthermore, the third and fourth key questions showed that after moving to Switzerland, reduced exposure to English, a desire to make friends, ongoing stress, anxiety, boredom, and a lack of motivation and school engagement, most likely caused English language inhibition and a gradual proficiency decline, leading me to forget the language I once spoken fluently as a young boy in the USA.

Taking a transdisciplinary perspective and drawing on the interconnected areas of mind, brain, health, and education (i.e., MBHE Science) has provided several potential answers and explanations for the acquisition and retrieval of my spoken English. Before ending the paper, I will now briefly discuss two specific pedagogical implications: (1) the consideration of students' emotional states, and (2) the integration of new information into students' existing knowledge of language. These implications, like the other parts of this paper, are subjective and derived from my own unique experiences; nonetheless, I believe they hold pedagogical value and contribute to the discussion of enhancing students' language learning process in the classroom.

Pedagogical implications

As discussed in the fourth key question, I was often stressed, anxious, and bored in middle school, with emotional triggers, such as stress, anxiety, and boredom in the classroom, impeding learning and impacting memory. Rodski (2019) argues that "[t]he basic cause of all stress is the perception that you can't control a situation the way you want to" (p. 107) and when "confronted with relentless demands and unexpected challenges, we tend to slip into negative emotions" (p. 100) that impede learning. However, low to mild levels of stress – called 'eustress' (Kloidt & Barsalou, 2024) – activate the amygdala, prompt the hippocampus to consolidate memory, and stimulate alertness. It is, therefore, paramount for teachers to consider the emotional state of their students and maintain an appropriate level of challenge and stress. Thus, "[k]eeping students in a neuroplastic 'sweet spot' of arousal is a core element in the art of teaching" (Cozolino, 2013, pp. 81-82). If this is achieved, students will feel safe in the L2 classroom and be more likely to take risks and experiment with new language, thereby enhancing their language learning process in a secure and positive classroom environment.

In considering students' emotional states, teachers should strive to create an enjoyable L2 classroom environment, irrespective of their students' age and proficiency levels. This may include the occasional use of fun and engaging class activities that require little cognitive demand, such as watching a funny movie clip or having students share something about their own culture. L2 teachers should also actively promote the notion that it is never too late to learn an additional language because "our brain is plastic and capable of learning throughout the lifespan" (Nouri, 2015, p. 41). This reduces the pressure on students, especially when the value in the classroom is placed on the use of two or more languages, and if L2 learning is embraced as a complex and multifaceted process rather than a race toward the unrealistic goal of attaining native speaker proficiency in the target language (van Hell, 2023). Creating an emotionally safe, trusting, and enjoyable classroom atmosphere is paramount to enhance learning, but it is equally important for L2 teachers, given the emotionally stressful nature of language teaching (Gkonou et al., 2020).

Also relevant is that, even though I gradually forgot English in Switzerland, my neural pathways of the English language had been established in the USA. Prior knowledge of language (or any other knowledge) connects with and facilitates the integration of newly learned information into existing knowledge, which, in turn, improves the efficiency of learning (McDermott & Zerr, 2019). My experience

supports Chakraborty and Esposito's (2024) research, suggesting that connecting new language learning with previously learned or known language facilitates the reactivation – or retrieval – of this new language; in my case, English. In terms of classroom application, this means that language teachers would be well-advised to connect newly introduced language features to language their students have already studied or are familiar with (Lindsay & Gaskell, 2010). This could, for instance, entail a comparison of grammatical similarities between two languages, such as Swiss-German and English. Comparing language structures facilitates positive crosslinguistic transfer (van Hell, 2023) and thus augments learning. Or, teachers could introduce a new vowel sound and then link it to a previously learned sound or a specific one that students have in their first language, further illustrating the similarities between the two languages. Making such connections, ideally in combination with the consideration of learners' backgrounds and maintenance of an appropriate level of stress, most likely enhances the language learning process, especially between two languages that belong to the same language family, such as English and Swiss-German.

Finally, it is worth noting that the relationship between instructional methods and brain activity during language learning is not well understood (van Hell, 2023). Nonetheless, considering students' emotions and the connection between new content and existing knowledge are two practical ideas, although several other implications could, of course, be derived from a language journey such as the one described and analysed in this paper. I hope this paper prompts readers to consider some additional pedagogical suggestions.

Conclusion

Exploring the four key questions has shed some light on the acquisition and retrieval process of my spoken English. Yet, learning is a dynamic and personal process (Immordino-Yang & Gotlieb, 2018). The formation of neural pathways depends on the immediate environment in conjunction with previous experiences (Tokuhama-Espinosa, 2019), with individual differences, including language aptitude and age, playing key roles in the language learning process, memory formation, and storage of each individual learner (e.g., Li et al., 2014; Suzukida, 2021; Turker et al., 2021). As such, the selection and discussion of the four key questions and their pedagogical implications were a subjective undertaking based on my own lived experiences. The various aspects discussed in this article have not been tested empirically; in other words, their transferability to other learners in similar contexts is somewhat limited.

The relationship between the mind, culture, and language (e.g., Shaules, 2019) could also have been investigated to provide additional insight into my acquisition and retrieval process. This is an important area to examine, given Kitayama and colleagues' (2019) suggestion that culture influences our genetic structure. Writing this paper has, therefore, led me to ponder the composition of a follow-up article that focuses specifically on the impact of the immersion environment on my English language retrieval and competence in NZ. For now, I hope that taking a transdisciplinary perspective and engaging in a self-exploration of learning, forgetting, and eventually retrieving my spoken English has provided some new and meaningful insights into how bilinguals access and process language, thereby opening the floor for further discussion in this area.

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to Professor Tracey Tokuhama-Espinosa and her teaching team. Their guidance and feedback were invaluable in shaping my thinking and helping me write this paper.

AUTHOR

Michael Burri is an Associate Professor in the Department of Curriculum and Instruction at Wenzhou-Kean University, and an Honorary Fellow in the School of Education at University of Wollongong. Prior to moving China, Michael taught/conducted research in Australia, Canada, and Japan. His professional interests include pronunciation instruction, teacher education/learning, Mind Brain Education, and the ethical use of GenAI. He has published in many leading journals, and his longitudinal research on learning to teach English pronunciation was awarded the TESOL Award for an Outstanding Paper on NNEST Issues (2015) and the MAK Halliday Prize for Outstanding Research in Applied Linguistics (2019).

REFERENCES

Barlow, F. K. (2019). Nature vs. nurture is nonsense: On the necessity of an integrated genetic, social, developmental, and personality psychology. *Australian Journal of Psychology, 71*(1), 68–79. https://doi.org/10.1111/ajpy.12240

Betts, K., Miller, M., Tokuhama-Espinosa, T., Shewokis, P., Anderson, A., Borja, C., Galoyan, T., Delaney, B., Eigenauer, J., & Dekker, S. (2019). *International report: Neuromyths and evidence-based practices in higher education*. Online Learning Consortium.

Bergelson, E., Soderstrom, M., Schwarz, I.-C., Rowland, C. F., Ramírez-Esparza, N., R. Hamrick, L., Marklund, E., Kalashnikova, M., Guez, A., Casillas, M., Benetti, L., Alphen, P. v., & Cristia, A. (2023). Everyday language input and production in 1,001 children from six continents. *Proceedings of the National Academy of Sciences, 120*(52), e2300671120. https://doi.org/doi:10.1073/pnas.2300671120

Bergen, B. K. (2012). *Louder than words: The new science of how the mind makes meaning*. Basic Books.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition, 12*(1), 3–11. https://doi.org/10.1017/S1366728908003477

Bloom, M. S., Orthmann-Murphy, J., & Grinspan, J. B. (2022). Motor learning and physical exercise in adaptive myelination and remyelination. *ASN Neuro, 14*. https://doi.org/10.1177/17590914221097510

Broad, D. (2020). Literature review of theories of second language acquisition. *Journal of Applied Linguistics and Language Research, 7*(1), 80–86.

Burri, M. (2022). From ESL student to teacher educator: Reflections on transnational and transcultural professional identity development. In G. Barkhuizen (Ed.), *Language teachers studying abroad: Identities, emotions and disruptions* (245–255). Multilingual Matters.

Burri, M. (2023, July). *L2 teaching and learning: What does neuroscience have to do with this?*. Plenary given at the 4th Mekong TESOL Conference, Can Tho City, Vietnam.

Cantor, P. (2021). All children thriving: A new purpose for education. *American Educator, 45*(3), 14–26. Available at https://files.eric.ed.gov/fulltext/EJ1322300.pdf

Casado, A., Szewczyk, J., Wolna, A., & Wodniecka, Z. (2022). The relative balance between languages predicts the degree of engagement of global language control. *Cognition, 226*, 105169. https://doi.org/10.1016/j.cognition.2022.105169

Casanave, C. P. (2012). Diary of a dabbler: Ecological influences on an EFL teacher's efforts to study Japanese informally. *TESOL Quarterly*, *46*(4), 642–670. https://doi.org/10.1002/tesq.47

Chakraborty, J., & Esposito, A. G. (2024). Adult learners self-derive new knowledge through integration of novel information and prior knowledge and are more successful with reactivation. *Mind, Brain, and Education, 18*(3), 226–235. https://doi.org/10.1111/mbe.12409

Chaouch-Orozco, A., & Martín-Villena, F. (2025). Network science reveals the early signs of L1 lexical attrition: Introducing the Lexical Attrition Foundation (LeAF) framework. *Bilingualism: Language and Cognition*, *28*(1), 43–53. https://doi.org/10.1017/S1366728924000063

Chomsky, N. (1965). *Aspects of the theory of syntax*. The MIT Press.

Chomsky, N. (2015). *Aspects of the theory of syntax: 50th anniversary edition*. The MIT Press.

Cohen, A. D. (Ed.). (1980). *The neurolinguistics of second language learning*. Available at https://files.eric.ed.gov/fulltext/ED202252.pdf

Cook, V. (2016). *Second language learning and language teaching* (5th ed.). Routledge.

Costandi, M. (2016). *Neuroplasticity*. The MIT Press.

Cozolino, L. (2013). *The social neuroscience of education: Optimizing attachment and learning in the classroom*. W. W. Norton & Company, Inc.

Cummins, J. (1999). *BICS and CALP: Clarifying the distinction* (ERIC Document Reproduction Service No. ED438551).

Damasio, A. (2021). *Feeling & knowing: Making minds conscious*. Robinson.

Dolcos, F., Katsumi, Y., Moore, M., Berggren, N., de Gelder, B., Derakshan, N., Hamm, A. O., Koster, E. H. W., Ladoucer, C. D., Okon-Singer, H., Pegna, A. J., Richter, T., Schweizer, S., Van den Stock, J., Ventura-Bort, C., Weymar, M., & Dolcos, S. (2020). Neural correlates of emotion-attention interactions: From perception, learning, and memory to social cognition, individual differences, and training interventions. *Neuroscience & Biobehavioral Reviews, 108*, 559–601. https://doi.org/10.1016/j.neubiorev.2019.08.017

Duff, P. A., & May, S. (Eds.) (2017). *Language socialisation* (3rd ed.). Springer.

Eising, E., Mirza-Schreiber, N., de Zeeuw, E. L., Wang, C. A., Truong, D. T., Allegrini, A. G., Shapland, C. Y., Zhu, G., Wigg, K. G., Gerritse, M. L., Molz, B., Alagöz, G., Gialluisi, A., Abbondanza, F., Rimfeld, K., van Donkelaar, M., Liao, Z., Jansen, P. R., Andlauer, T. F. M., . . . Fisher, S. E. (2022). Genome-wide analyses of individual differences in quantitatively assessed reading- and language-related skills in up to 34,000 people. *Proceedings of the National Academy of Sciences, 119*(35), e2202764119. https://doi.org/doi:10.1073/pnas.2202764119

Fabbro, F. (1999). *The neurolinguistics of bilingualism: An introduction*. Psychology Press.

Farid, A., Elbakai, F., & Fanani, A. (2021). Children's innate capacity of learning the first language: An overview of structure-dependent rules. *Journal of Research in Foreign Language Teaching, 2*(1), 1–8.

Filipović, J. (2019). Transdisciplinary qualitative paradigm in applied linguistics: Autoethnography, participatory action research and minority language teaching and learning. *International Journal of Qualitative Studies in Education, 32*(5), 493-509. https://doi.org/10.1080/09518398.2019.1597209

Fox, S. E., Levitt, P., & Nelson III, C. A. (2010). How the timing and quality of early experiences influence the development of brain architecture. *Child Development, 81*(1), 28–40. https://doi.org/10.1111%2Fj.1467-8624.2009.01380.x

Friederici, A. D. (2017). *Language in our brain: The origins of a uniquely human capacity*. The MIT Press.

Garton, S., & Copland, F. (Eds.) (2022). *The Routledge handbook of teaching English to young learners*. Routledge.

Gershman, S. J. (2023). The molecular memory code and synaptic plasticity: A synthesis. *Biosystems, 224*, 104825. https://doi.org/10.1016/j.biosystems.2022.104825

Gkonou, C., Dewaele, J.-M., & King, J. (Eds.). (2020). *The emotional rollercoaster of language teaching*. Multilingual Matters.

Hartas, D. (2011). Families' social backgrounds matter: Socio-economic factors, home learning and young children's language, literacy and social outcomes. *British Educational Research Journal, 37*(6), 893–914. http://dx.doi.org/10.1080/01411926.2010.506945

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. Wiley.

Hiver, P., Obando, G., Sang, Y., Tahmouresi, S., Zhou, A., & Zhou, Y. (2019). Reframing the L2 learning experience as narrative reconstructions of classroom learning. *Studies in Second Language Learning and Teaching, 9*(1), 83–116. https://doi.org/10.14746/ssllt.2019.9.1.5

Hoff, E., & Core, C. (2015). What clinicians need to know about bilingual development. *Seminars in Speech and Language, 36*(2), 89–99.

https://doi.org/10.1055/s-0035-1549104

Immordino-Yang, M. H., & Gotlieb, R. J. M. (2018). An evolving understanding of social emotions from a mind, brain, and education perspective. In M. S. Schwartz & E. J. Paré-Blagoev (Eds.), *Research in mind, brain, and education* (pp. 73–96). Routledge.

Immordino-Yang, M. H., & Damasio, A. (2007). We feel, therefore we learn: The relevance of effective and social neuroscience to education. *Mind, Brain, and Education, 1*(1), 3–10. https://doi.org/10.1111/j.1751-228X.2007.00004.x

Immordino-Yang, M. H., Darling-Hammond, L., & Krone, C. R. (2019). Nurturing nature: How brain development is inherently social and emotional, and what

this means for education. *Educational Psychologist, 54*(3), 185–204. https://doi.org/10.1080/00461520.2019.1633924

Kail, M., & Isel, F. (2021). Language, plasticity, and learning: Challenges at the forefront of research. *Language, Interaction and Acquisition, 12*(1), 1–9. https://doi.org/10.1075/lia.00012.int

Kitayama, S., Varnum, M. E. W., & Salvador, C. E. (2019). Cultural neuroscience. In D. Cohen & S. Kitayama (Eds.), *Handbook of cultural psychology* (2nd ed.) (pp. 79–118). The Guilford Press.

Kloidt, J., & Barsalou, L. W. (2024). Establishing a Comprehensive Hierarchical construct of Eustress (CHE). *Current Psychology, 43,* 32258–32273. https://doi.org/10.1007/s12144-024-06750-7

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*(2), F13–F21. https://doi.org/10.1111/j.1467-7687.2006.00468.x

Kurkela, J. L. O., Hämäläinen, J. A., Leppänen, P. H. T., Shu, H., & Astikainen, P. (2019). Passive exposure to speech sounds modifies change detection brain responses in adults. *NeuroImage, 188*, 208–216. https://doi.org/https://doi.org/10.1016/j.neuroimage.2018.12.010

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press.

Li, P., Legault, J., & Litcofsky, K. A. (2014). Neuroplasticity as a function of second language

learning: Anatomical changes in the human brain. *Cortex, 58*, 301–324. https://doi.org/10.1016/j.cortex.2014.05.001

Li, J., Yao, C., Li, Y., Liu, X., Zhao, Z., Shang, Y., Yang, J., Yao, Z., Sheng, Y., & Hu, B. (2024). Effects of second language acquisition on brain functional networks at different developmental stages. *Brain Imaging and Behavior, 18*, 808–818. https://doi.org/10.1007/s11682-024-00865-y

Lindquist, K. A., MacCormack, J. K., & Shablack, H. (2015). The role of language in emotion: predictions from psychological constructionism. *Frontiers in Psychology, 6*, 444. https://doi.org/10.3389/fpsyg.2015.00444

Lindsay, S., & Gaskell, M.G. (2010). A complementary systems account of word learning in L1 and L2: Learning words in L1 and L2. *Language Learning, 60*(s2), 45–63.

https://doi.org/10.1111/j.1467-9922.2010.00600.x

Long, M. H. (1981). Input, interaction, and second language acquisition. *Annals of the New York Academy of Sciences, 379*(1), 259–278.

https://doi.org/10.1111/j.1749-6632.1981.tb42014.x

Kaller, M. S., Lazari, A., Blanco-Duque, C., Sampaio-Baptista, C., & Johansen-Berg, H. (2017). Myelin plasticity and behaviour—connecting the dots. *Current Opinion in Neurobiology, 47*, 86–92. https://doi.org/10.1016/j.conb.2017.09.014

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.

McDermott, K. B., & Zerr, C. L. (2019). Individual differences in learning efficiency. *Current Directions in Psychological Science, 28*(6), 607–613. https://doi.org/10.1177/0963721419869005

Morgan-Short, K., & van Hell, J. G. (Eds.). (2023). *The Routledge handbook of second language acquisition and neurolinguistics*. Taylor & Francis.

Murray, B. D., Holland, A. C., & Kensinger, E. A. (2013). Episodic memory and emotion. In M. D. Robinson, E. R. Watkins, & E. Harmon-Jones (Eds.), *Handbook of cognition and emotion* (pp. 156–175). The Guilford Press.

Neely, S. A., Williamson, J. M., Klingseisen, A. Zoupi, L., Early, J. J., Williams, A., & Lyons, D. A. (2022). New oligodendrocytes exhibit more abundant and accurate myelin regeneration than those that survive demyelination. *Nature Neuroscience, 25*, 415–420. https://doi.org/10.1038/s41593-021-01009-x

Nouri, A. (2015). Cognitive neuroscience of foreign language education: Myths and realities.

Iranian Journal of Research in English Language Teaching, 3(1), 40-47.

Oppenheim, G. M., Griffin, Z., Peña, E. D., & Bedore, L. M. (2020). Longitudinal evidence for simultaneous bilingual language development with shifting language dominance, and how to explain it. *Language Learning, 70*(S2), 20–44. https://onlinelibrary.wiley.com/doi/abs/10.1111/lang.12398

Ortuño-Sahagún, D., Schliebs, R., & Pallàs, M. (2019). Editorial: Epigenetic mechanisms regulating neural plasticity. *Frontiers in Cellular Neuroscience, 13*, 118. https://doi.org/10.3389/fncel.2019.00118

Patel, M., Solly, M., & Copeland, S. (2022). *The future of English: Global perspectives*. British Council.

Pawlak, M., Kruk, M., & Zawodniak, J. (2024). *Teachers reflecting on boredom in the language classroom*. Equinox.

Perrin, D., & Kramsch, C. (Eds.) (2019). Transdisciplinarity in applied linguistics. *AILA Review, 31*(1). https://doi.org/10.1075/aila.31

Persici, V., Vihman, M., Burro, R., & Majorano, M. (2019). Lexical access and competition in bilingual children: The role of proficiency and the lexical similarity of the two languages. *Journal of Experimental Child Psychology, 179*, 103–125. https://doi.org/10.1016/j.jecp.2018.10.002

Pliatsikas, C., DeLuca, V., & Vois T., (2020). The many shades of bilingualism: Language experiences modulate adaptations in brain structure. *Language Learning, 70*(S2), 133–149. https://doi.org/10.1111/lang.12386

Power, J. D., & Schlaggar, B. L. (2017). Neural plasticity across the lifespan. *Wiley Interdisciplinary Reviews. Developmental Biology, 6*(1), 10.1002/wdev.216. https://doi.org/10.1002/wdev.216

Quan, L., Uyeda, A., & Muramatsu, R. (2022). Central nervous system regeneration: The roles of glial cells in the potential molecular mechanism underlying remyelination. *Inflammation and Regeneration, 42*(7), 1–12. https://doi.org/10.1186/s41232-022-00193-y

Saito, K., Sun, H., Kachlicka, M., Robert, J., Nakata, N., & Tierney, A. (2022). Domain-general auditory processing explains multiple dimensions of L2 acquisition in adulthood. *Studies in Second Language Acquisition, 44*(1), 57–86. https://doi.org/10.1017/S0272263120000467

Saito, K., Suzukida, Y., Tran, M., & Tierney, A. (2021). Domain-general auditory processing partially explains L2 speech learning in classroom settings: A review and generalization study. *Language Learning, 71*(3), 669–715. https://doi.org/10.1111/lang.12447

Salsano, I., Tain, R., Giulietti, G., Williams, D. P., Ottaviani, C., Antonucci, G., Thayer, J. F., & Santangelo, V. (2024). Negative emotions enhance memory-guided attention in a visual search task by increasing frontoparietal, insular, and parahippocampal cortical activity. *Cortex, 173*, 16–33. https://doi.org/10.1016/j.cortex.2023.12.014

Schmid, M. S., & Köpke, B. (Eds.). (2019). *The Oxford handbook of language attrition*. Oxford University Press.

Schwieter, J. W., & Rastelli, S. (2024). Neurolinguistics in language learning and teaching. In L. Wei, Z. Hua, & J. Simpson (Eds.) *The Routledge handbook of applied linguistics* (pp. 362–373). Routledge.

Shao, K., Nicholson, L. J., Kutuk, G., & Lei, F. (2020). Emotions and instructed language learning: Proposing a second language emotions and positive psychology model. *Frontiers in Psychology, 11*, 2142. https://doi.org/10.3389/fpsyg.2020.02142

Shaules, J. (2019). *Language, culture, and the embodied mind: A developmental model of linguaculture learning*. Springer

Sudarshan, A., & Baum, S. R. (2019). Bilingual lexical access: A dynamic operation modulated by word-status and individual differences in inhibitory control. *Bilingualism: Language and Cognition, 22*(3), 537–554. https://doi.org/10.1017/S1366728918000111

Suzuki, Y., Jeong, H., Cui, H., Okamoto, K., Kawashima, R., & Sugiura, M. (2023). An fMRI validation study of the word-monitoring task as a measure of implicit knowledge: Exploring the role of explicit and implicit aptitudes in behavioral

and neural processing. *Studies in Second Language Acquisition, 45*(1), 109–136. https://doi.org/10.1017/S0272263122000043

Suzukida, Y. (2021). The contribution of individual differences to L2 pronunciation learning: Insights from research and pedagogical implications. *RELC Journal, 52*(1), 48-61. https://doi.org/10.1177/0033688220987655

Thomas, M. S. C., Kovas, Y., Meaburn, E., & Tolmie, A. (2015). What can the study of genetics offer to educators? *Mind, Brain and Education, 9*(2), 72–80. https://doi.org/10.1111/mbe.12077

Tokuhama-Espinosa, T. (2008). *Living languages: Multilingualism across the lifespan*. Praeger Publishers.

Tokuhama-Espinosa, T. (2018). *Neuromyths: Debunking false ideas about the brain*. W. W. Norton & Company Ltd.

Tokuhama-Espinosa, T. (2019). *Five pillars of the mind: Redesigning education to suit the brain*. W. W. Norton & Company.

Tokuhama-Espinosa, T., Simmers, K., Batchelor, D., Nelson, A. D., & Borja, C. (2023). A theory of mental frameworks. *Frontiers in Psychology, 14*(1220664). https://doi.org/10.3389/fpsyg.2023.1220664

Tomassy, G. S., Dershowitz, L. B., & Arlotta, P. (2016). Diversity matters: A revised guide to myelination. *Trends in Cell Biology, 26*(2), 135–147. https://doi.org/10.1016/j.tcb.2015.09.002

Trammell, J. P., & Clore, G. L. (2014). Does stress enhance or impair memory consolidation? *Cognition & Emotion, 28*(2), 361–374. https://dx.doi.org/10.1080%2F02699931.2013.822346

Tu, L., Wang, J., Abutalebi, J., Jiang, B., Pan, X., Li, M., Gao, W., Yang, Y., Liang, B., Lu, Z., & Huang, R. (2015). Language exposure induced neuroplasticity in the bilingual brain: A follow-up fMRI study. *Cortex, 64*, 8-19. https://doi.org/10.1016/j.cortex.2014.09.019

Turker, S., Seither-Preisler, A., & Reiterer, S. M. (2021). Examining individual differences in language learning: A neurocognitive model of language

aptitude. *Neurobiology of Language, 2*(3), 389–415. https://doi.org/10.1162/nol_a_00042

Tyng, C. M., Amin, H. U., Saad, M. N., & Malik, A. S. (2017). The influences of emotion on learning and memory. *Frontiers in Psychology, 8*, 1454. https://doi.org/10.3389/fpsyg.2017.01454

Wang, F., Yang, J., Pan, F., Ho, R. C., & Huang, J. H. (2020). *Editorial: Neurotransmitters and emotions*. *Frontiers in Psychology, 11*(21), 1–3. https://doi.org/10.3389/fpsyg.2020.00021

Wei, M., Joshi, A. A., Zhang, M., Mei, L., Manis, F. R., He, Q., Beattie, R. L., Xue, G., Shattuck, D. W., Leahy, R. M., Xue, F., Houston, S. M., Chen, C., Dong, Q., & Lu, Z. L. (2015). How age of acquisition influences brain architecture in bilinguals. *Journal of Neurolinguistics*, *36,* 35–55. https://doi.org/10.1016/j.jneuroling.2015.05.001

Wessels, S., & Trainin, G. (2023). From bilingual to biliteracy: Learning from families. *Early Childhood Education Journal, 53*, 315–326. https://doi.org/10.1007/s10643-023-01519-2

Willis, J., & Willis, M. (2020). *Research-based strategies to ignite learning: Insights from neuroscience and the classroom* (2nd ed.). ASCD.

Willmore, B. D., & King, A. J. (2023). Adaptation in auditory processing. *Physiological Reviews, 103*(2), 1025–1058. https://doi.org/10.1152/physrev.00011.2022

Wodniecka, Z., Casado, A., Kałamała, P., Marecka, M., Timmer, K., & Wolna, A. (2022). The dynamics of language experience and how it affects language and cognition. In K. D. Federmeier & H-W. Huang (Eds.), *The psychology of learning and motivation, volume 72: Adult and second language learning* (pp. 235-281). Elsevier. https://doi.org/10.1016/bs.plm.2020.02.005

van Hell, J. G. (2023). The Neurocognitive underpinnings of second language processing: Knowledge gains from the past and future outlook. *Language Learning*, *73*(S2), 95–138 https://doi.org/10.1111/lang.12601

Zappa, A., Bolger, D., Pergandi, J-M., Fargier, R., Mestre, D., & Frenck-Mestre, C. (2024). The neural correlates of embodied L2 learning: Does embodied L2

verb learning affect representation and retention?. *Neurobiology of Language, 5*(2), 360–384. https://doi.org/10.1162/nol_a_00132

Zein, S., & Butler, Y. G. (Eds.) (2024). *English for young learners in Asia: Challenges and directions for teacher education*. Routledge.

Vietnamese EFL University Students' Perceptions and Experiences of Using the ELSA Speak Application for Developing English Oral Fluency

Diem Tran Thi Ngoc[1]

University of Sydney

ABSTRACT

This case study examines Vietnamese EFL university students' perceptions and experiences with using the ELSA Speak application to enhance their English oral fluency. Over a four-week period, seven participants engaged with the app, having data collected through screen recordings, daily reflective diaries and semi-structured interviews. Thematic analysis revealed that learners valued a combination of authentic interaction, structured repetition and explicit feedback. Features that supported personalisation, motivation, and future goals, such as 'Study by Topic', 'Improve Pronunciation', 'ELSA AI Conversation', 'Practice Daily Lessons', and 'Earn a Certificate', were especially valued. However, technical limitations, content lacking depth of expertise and inadequate tracking features emerged as sources of frustration that could impede sustained engagement.

Keywords: AI-assisted language learning, artificial intelligence (AI), ELSA Speak, mobile-assisted language learning, skill acquisition theory

[1] Correspondence: Diem Tran Thi Ngoc: Email: ngocdiem.tranthi@gmail.com; https://orcid.org/0009-0000-8309-5057

ISSN 1834-4712 (Online).

DOI: 10.5281/zenodo.17588104

Citation: Tran, D. T. N. (2025). Vietnamese EFL university students' perceptions and experiences of using the ELSA Speak application for developing English oral fluency. *University of Sydney Journal of TESOL*, *4*, 33–83. https://doi.org/10.5281/zenodo.17588104

INTRODUCTION

As English has become a crucial means of international communication, proficiency in spoken English has been a significant goal for many learners, particularly in countries where English is considered a foreign language (Kirkpatrick, 2010). Among all the speaking skills, the quest for oral fluency in the English as a Foreign Language (EFL) setting has gained increasing significance, especially within the context of globalisation. However, despite years of classroom instruction, many EFL students may be able to produce grammatically correct sentences on paper but struggle to speak fluently and confidently (Farrokhi & Mahmoudi, 2014).

In Vietnam, developing learners' English-speaking proficiency is often neglected in test-oriented education (Thao & Xuan, 2020). Many high school graduates are fearful of making mistakes and not confident in speaking English (Van Tuyen & Loan, 2019). In 2016, the Ministry of Education and Training reported that the National Foreign Languages Project 2020, which received nearly 500 million dollars in funding to improve the country's English education, had not met its targets, and the failure was due mainly to unrealistic benchmarks and insufficient instructional time (Tran & Tanemura, 2020).

In response to these challenges, Mobile-Assisted Language Learning (MALL) has emerged as a promising alternative. The rapid growth of smartphones and mobile applications has transformed the way learners access language resources. Among the range of mobile applications available, those powered by generative Artificial Intelligence (AI)—such as ELSA Speak—have gained notable attention. Evers and Chen (2020) noted that these apps employ advanced speech recognition technologies to provide instant feedback on pronunciation, fluency and intonation and offer learners a sense of autonomy and control over their language learning.

Despite the growing popularity of ELSA Speak (Kaiser, 2024), limited research has examined in-depth learners' experiences with the app, particularly in the Vietnamese EFL context. Most existing studies are experimental, focusing on quantitative outcomes such as test scores or pronunciation accuracy, whereas few have explored how learners perceive, interact with, and evaluate the app (discussed further below). Meanwhile, understanding these user experiences is vital for evaluating the app's pedagogical values, identifying its strengths and limitations, and informing future design and implementation. Therefore, the current study aims to address this gap by examining Vietnamese EFL university

students' perceptions and experiences with ELSA Speak for the development of their oral fluency.

REVIEW OF PREVIOUS RESEARCH

This section reviews previous studies on Mobile-Assisted Language Learning (MALL), focusing on its application in oral fluency development in EFL contexts and learners' perceptions of using mobile applications in their language learning.

MALL and oral fluency in EFL contexts

MALL involves the use of mobile devices to support second or foreign language learning. In fact, Melhuish and Falloon (2010) highlight five defining features of MALL—portability, ubiquity, situated learning, connectivity, and personalisation—which allow learners to access language input beyond the classroom. Recent years have seen significant advances in integrating artificial intelligence (AI) into MALL, particularly in speaking practice and assessment (Zou et al., 2023). Unlike traditional technologies, AI-integrated systems provide real-time, personalised feedback, offering learners greater flexibility, interactive support and unlimited practice opportunities that may not be feasible with human instructors (Fathi & Rahimi, 2024; Kuddus, 2022).

Additionally, advances in machine learning, Natural Language Processing (NLP), and Automatic Speech Recognition (ASR) have increased the potential for even more tailored language learning experiences. MALL applications now provide learners with authentic, socially-situated speaking practice through interactive environments (Tapalova & Zhiyenbayeva, 2022). For instance, conversational AI agents, such as Google Assistant, can engage learners in spoken interactions and offer scaffolded feedback in low-pressure settings, compared to conventional classrooms (Chen et al., 2023; Istrate, 2018).

Specifically, in EFL contexts—where exposure to spoken English is often limited—MALL has shown promise in enhancing oral fluency. Empirical studies support this claim. Grimshaw and Cardoso (2018), for example, found that learners using the mobile game Spaceteam ESL demonstrated greater fluency gains than those in the control group. Similarly, Ahmed et al. (2022) reported significant improvements in speaking fluency among EFL learners who used Duolingo and WhatsApp, attributing these gains to the frequent, anxiety-free speaking opportunities

provided. These findings suggest that MALL facilitates the proceduralisation of language, a key factor in developing fluency.

However, a review of the literature reveals a predominant focus on structured tasks (e.g., read-aloud), with limited research on AI use in spontaneous speech involving open-ended responses (Zechner & Hsieh, 2024). In this regard, Rudnik (2024) distinguishes "traditional AI," which operates according to preset rules, from "generative AI," which can produce original content and support adaptive interaction. As generative tools become more widespread, users may mistakenly assume that all AI technologies possess such capabilities, despite many lacking the necessary functionality for spontaneous speech (Kaiser, 2024).

Perceptions towards MALL

Research has shown that learner perceptions play a crucial role in sustained engagement with technology for language learning (Gardner & Lambert, 1972). Overall, the literature reveals a generally positive attitude towards MALL. For instance, Van and Duong (2022) found that Vietnamese university students perceived MALL as useful, particularly for vocabulary and pronunciation, and found it easy to use. Similarly, Zou et al. (2023) reported that learners found AI-based speech evaluation tools to be enjoyable and helpful. However, they raised concerns about the potential for inaccurate feedback and a lack of human interaction.

In another study, Hoi and Mu (2021) investigated Vietnamese EFL students' expectations of teachers' roles in promoting MALL. Analysing survey data from 293 university learners using a Rasch-based path model, the study found that students valued teacher guidance on using mobile resources both in and beyond the classroom, suggesting that the successful implementation of MALL depends not only on technological appeal but also on pedagogical instruction.

Besides the positive feedback, there have been accounts of shortcomings in mobile applications. Learners reported dissatisfaction, mainly due to overly complex or technical feedback and poor interface design (An et al., 2023; Wang et al., 2024). Another notable recurring issue is the difficulty in interacting with voice recognition systems, which frequently leads to frustration and disengagement (Dizon et al., 2022; McCrocklin, 2019).

Reviewing ELSA Speak, Loora, and Vocal Image applications, Kaiser (2024) noted that although features such as ASR and generative AI may enhance user

engagement, persistent challenges remain, particularly in providing accurate suprasegmental feedback and processing natural speech. ELSA Speak, for example, may support communicative competence but often misjudges connected speech and relies on simplistic indicators, such as loudness, to assess stress and rhythm, reflecting limited input from linguistics experts. Moreover, AI speech evaluation systems were reported to fall short of human listeners in recognising different accents and spontaneous speech, raising concerns about demotivating learners and reducing their confidence (Evers & Chen, 2020; Jeon, 2022; McCrocklin, 2019; Zou et al., 2023).

Research gaps and research questions

The review of the literature has shown that while several prior studies has explored the impact of MALL tools, including ELSA Speak, on English language learning, particularly pronunciation and speaking skills (e.g., Bashori et al., 2024; Ebadi & Ebadijalal, 2022; Rad, 2024), most studies are experimental or mixed methods, with limited focus on oral fluency specifically (Tavakoli & Hunter, 2018). Furthermore, little is known about the in-depth perceptions of Vietnamese EFL university students regarding ELSA Speak and the perceived usefulness of its individual features for developing fluency. Therefore, to address this gap, the present study adopts a qualitative case study approach to investigate these learners' experiences and perceptions, thereby contributing insights to the emerging field of AI-enhanced fluency development in MALL. The following two research questions were formulated to address the identified research gaps:

1. What are the valuable features of ELSA Speak to improve English oral fluency perceived by Vietnamese EFL university students?

2. What are Vietnamese EFL university students' general perceptions towards the advantages and disadvantages of using ELSA Speak for improving fluency?

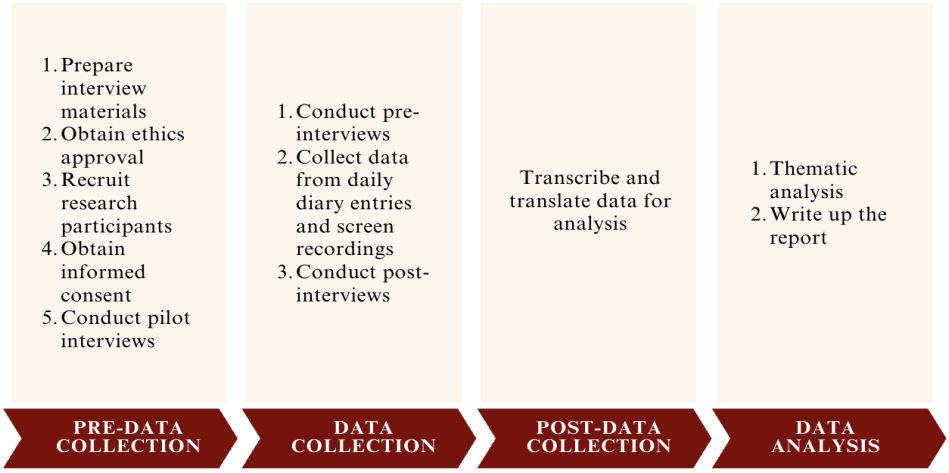
RESEARCH METHODOLOGY

The current study employed a qualitative case study approach to conduct an in-depth exploration of learners' perceptions and experiences within a bounded context (van Lier, 2005). According to Duff (2010), qualitative research enables researchers to gain an understanding of complex issues without the intention of manipulating circumstances or proving a pre-established hypothesis. For this

reason, the qualitative approach was suitable for the current study, as it enabled an in-depth investigation of learners' perceptions and experiences when engaging with ELSA Speak for the development of oral fluency. The study focused on a group of seven Vietnamese EFL university students who voluntarily used the ELSA Speak app over four weeks. This design enables a deeper understanding of how learners interact with the app in their daily usage.

The study was first reviewed and approved by the Human Research Ethics Committee of the University of Sydney (Approval No. 2024/HE001278). Since the current study involved human participants, ethical considerations were addressed before data collection commenced and throughout the study. This aligned with Fraenkel et al. (2019), who suggested three important issues to be addressed in research: protecting participants from harm, ensuring confidentiality of research data, and considering the ethical implications of deceiving participants. Figure 1 summarises the detailed procedures for data collection and analysis undertaken in this study.

Figure 1. Research procedures



The pre-data collection stage involved preparing interview materials, obtaining ethics approval, recruiting participants, obtaining informed consent and conducting pilot interviews. The data collection phase included pre-interviews, diary entry and screen recording collection, and post-interviews. As shown here, data were collected through three primary instruments: semi-structured

interviews, reflective learner diaries, and screen recordings of app usage. In the post-data collection phase, all data were transcribed and translated in preparation for analysis. The data analysis phase employed thematic analysis, guided by Braun and Clarke's (2006) six-phase framework. To enhance the trustworthiness of the findings, intra-coder reliability checks were conducted, and data triangulation was carried out across the three data sources.

Research settings

After approval by the Human Research Ethics Committee of the University of Sydney, a recruitment advertisement, presented in both English and Vietnamese, was disseminated across Facebook pages and forums to engage potential participants, particularly students from universities across Vietnam. As the researcher was based in Australia while the participants were in Vietnam, the study was conducted entirely online, utilising platforms such as Zoom and Zalo Messenger, a widely used local application predominant among Vietnamese people for communication. This online approach was designed to facilitate broader geographic reach and to include a more diverse range of participants from various majors and universities across Vietnam, potentially encompassing a wider range of accents, educational backgrounds, and language learning experiences, and offering richer insights into the investigated questions.

Research participants

Initially, eight EFL students submitted their signed Participant Consent Form. However, one participant withdrew due to personal reasons, resulting in a final sample of seven Vietnamese EFL university students (four females and three males) for the current study. Table 1 presents detailed information about the participants.

Table 1. *Participants' demography*

Name	**Gender**	**Grade**	**Major**	**Goals**
Participant 1	Male	Year 4	Information Technology (IT)	Effectively communicate in the workplace Improve TOEIC scores
Participant 2	Female	Year 2	Economics	Respond more quickly and fluently without hesitation
Participant 3	Male	Year 1	Marketing - Finance	Enter a higher-level IELTS preparation class
Participant 4	Female	Year 2	English Language	Fluent and confident English communication Improve IELTS scores
Participant 5	Female	Year 1	English Language	Increase English levels and be more confident in communication
Participant 6	Male	Year 2	Computing	Improve speaking skills for future interaction and negotiation at work Succeed in a global work environment Improve IELTS Speaking scores
Participant 8	Female	Year 3	Chinese Language	Achieve a targeted IELTS score Improve English pronunciation

As depicted in this table, the participant group encompassed a wide range of academic levels, ranging from first-year undergraduates to those in their final year. The diversity extended to their fields of study, with three participants majoring in Languages, two each in Economics, Marketing, and Finance, and an additional two in Computing and Information Technology. The participants' motivations for

learning English varied considerably although the primary focus was on enhancing oral proficiency. These can be categorised into short-term goals, such as achieving a specific test score, and long-term goals, including professional development and personal growth.

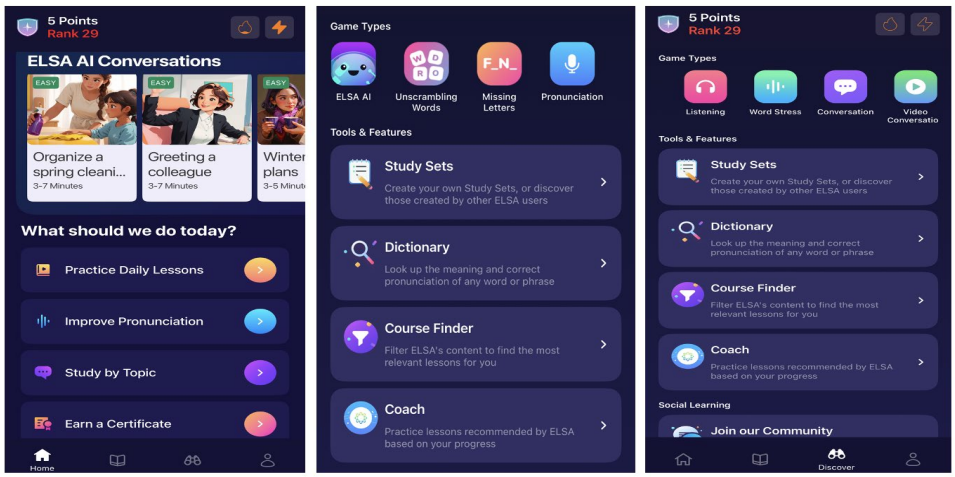
Research instruments and techniques

The current study employed four main instruments for data collection, which are presented in detail as follows.

The ELSA Speak application

The current study employed an AI-assisted application called ELSA Speak, specifically designed to support language learners' oral proficiency. The tool is designed to help users enhance their English-speaking skills, enabling them to communicate with clarity, fluency, and confidence (Karim et al., 2023). Figure 2 illustrates the ELSA Speak interface, highlighting its key features.

Figure 1. *Interface of ELSA Speak and its features*



As illustrated in Figure 2, ELSA Speak incorporates an extensive array of features, including 'ELSA AI Conversation', 'Practice Daily Lessons', 'Improve Pronunciation', 'Study by Topic', 'Earn a Certificate', 'Games and more'. According to ELSA Speak (n.d.), the application's standout functionalities encompass AI role-plays, an advanced AI coach, AI feedback and progress tracking, over 8,000 concise lessons, a personalised AI-guided learning pathway, a bilingual AI tutor, guided practice,

transcripts accompanied by vocabulary suggestions, and certificate courses, among others.

Given that the current study focuses on English oral fluency, the first research question aimed to explore which of these features participants perceived as most useful for improving their fluency proficiency. To address this, participants were asked to spend 20 to 30 minutes daily learning with ELSA Speak over four weeks, exploring the application's various functionalities. They were then asked to document their experiences, noting the usefulness of specific features and the application's advantages and disadvantages, in daily diaries and post-study interviews.

Semi-structured interviews

According to Yin (2015), semi-structured or open-ended interviews are a flexible yet focused form of qualitative interviewing in which researchers prepare a predetermined set of topics or questions, while allowing participants to guide the sequence and depth of discussion. This method is particularly suitable for the present study, as it enables participants to elaborate on their unique perspectives while allowing the researcher to gain a deeper understanding of their views on specific features, usage behaviours, and emotional responses. Besides that, its adaptability accommodates online formats and supports triangulation with other qualitative data sources (Seale, 1999). Noting the importance of a balanced exchange, Ezzy (2010) stressed that effective interviews should be shaped neither by the interviewer's dominance nor the interviewee's personal agenda. In line with this, the researcher made a conscious effort to build rapport, listen attentively, follow up, and elicit detailed responses. Following the interviews, audio recordings were transcribed and translated from Vietnamese to English and then analysed thematically.

Diaries

Given the aim of this study—to explore learners' perceptions and experiences as they engage with the AI-mediated ELSA Speak application—diary methods are particularly well-suited. Three key rationales underpin this appropriateness. Firstly, diary methods are a valuable data collection tool in applied linguistics and language learning research, adept at capturing participants' personal experiences, thoughts, and behaviours in real-time contexts (Porto, 2007). In this regard, Dörnyei (2007) noted that diaries enable participants to record language-related activities and reflect on their learning processes, providing access to their inner

thoughts on constructs such as motivation, self-regulation, anxiety, and autonomy, which can be challenging to capture through other methods or tools. Moreover, the reflective nature of diaries allows participants to articulate unique insights without the researcher needing to resort to indirect methods to access comparable thoughts (Dörnyei, 2007).

The diary method, while offering significant advantages, presented challenges that must be addressed. These included the substantial time and effort required from participants, which might result in incomplete or low-quality entries (Bolger et al., 2003), literacy demands, particularly for participants writing in a non-native language (Rose, 2020) and the potential for recall bias when entries are delayed (Carson & Longhini, 2002). To reduce the time and effort burden on participants, clear instructions, the expected time commitment (20-30 mins per day), guiding questions and introductory training were provided prior to data collection. To address literacy demands, participants were permitted to write in their native language, Vietnamese, thereby reducing linguistic barriers. As an example, an excerpt of Participant 2's writing in Vietnamese is presented in Figure 3.

Figure 3. Excerpt of participant 2's written reflections

Which activity do I find useful to improve my fluency? Why? (Tôi thấy hoạt động nào hữu ích để cải thiện khả năng nói trôi chảy của mình? Tại sao?)

Em thấy việc luyện nói shadowing về học tập, cuộc sống, sở thích,.. giúp em đỡ nhàm chán, hứng thú hơn với việc tự nói tiếng anh. Từ đó luyện được phản xạ và độ trôi chảy nói tiếng anh. Vì trong đoạn hội thoại có sẵn câu thoại trả lời, em biết thêm nhiều cách diễn đạt trong từng tình huống: như là khi đi ra sân bay kiểm tra hành lý nói như nào, tập sở thích mới trong dịch Covid,..

Luyện qua đoạn hội thoại thì mình vừa nghe được tông giọng, âm điệu như nào để bắt chước theo cho hợp lý.

Finally, to mitigate recall bias, as recommended by Ma and Oxford (2014), reminders were sent to participants to complete their diary entries immediately

after each practice session. These measures were designed to ensure the validity and reliability of the data collected through the diary method.

Screen recordings

According to Heycke and Spitzer (2019), screen recordings provide a powerful means of capturing real-time interactions in digital learning environments. This instrument enhances the transparency and reproducibility of research by preserving visual and procedural details that are often omitted from written reports. In the present study, screen recordings were used to document participants' multimodal interactions—including on-screen navigation, typing behaviour, and engagement with interface features—during their use of the ELSA Speak app. These forms of interaction would have been difficult to access through observations or interviews alone. Additionally, due to practical constraints, particularly the distance between the participants and the researcher, it was not feasible to observe the practice sessions in person. As such, screen recordings offered an alternative means of capturing detailed usage behaviour over time. The data collected from these recordings were later triangulated with diary entries to enhance the trustworthiness of the findings.

Despite their advantages, screen recordings also posed challenges, particularly regarding data privacy and the technical limitations associated with the large storage requirements of video files (Ho, 2021). To address these concerns, informed consent was obtained from all participants prior to data collection, and no video of participants' faces was recorded; only the mobile device screen during app use was captured. Furthermore, all data were securely stored in the University of Sydney's Research Data Storage system, with sufficient storage and access restricted solely to the researcher.

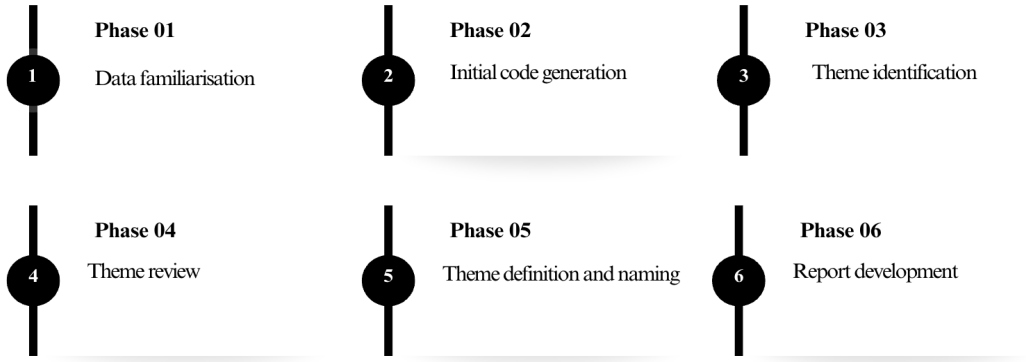
DATA ANALYSIS

Thematic analysis

To answer the two research questions, the current study employed thematic analysis (TA) as the primary method of data analysis. As defined by Braun and Clarke (2022), TA is a flexible yet systematic approach for identifying and interpreting meaningful patterns within a dataset, which enables researchers to analyse both manifest and latent meanings relevant to the research questions. TA can be positioned along a continuum between inductive and deductive approaches.

This was clarified by Braun and Clarke (2006). Whereas a deductive (top-down) approach applies pre-existing concepts to interpret data, an inductive (bottom-up) approach derives codes and themes directly from the data, prioritising participants' reported experiences. Given the exploratory nature of a case study, the current study primarily adopted an inductive approach to capture participants' authentic experiences and perceptions, free from predetermined analytical concepts. This analysis process, following Braun and Clarke's (2022) six-phase framework, is depicted in Figure 4.

Figure 4. *Thematic analysis stages (adopted from Braun and Clarke, 2022)*



As illustrated in Figure 4, the first TA phase involved the researcher getting thoroughly familiar with the data. Participants' interviews were transcribed and checked for accuracy. Both the interview transcripts and diary entries were then translated from Vietnamese into English. This process was relatively time-intensive, with the pre- and post-interview transcripts totalling 22,572 words and the diaries comprising 24,668 words. Following this, the researcher engaged in multiple readings of all transcripts and diaries to gain an overall impression of the dataset before coding, or, in other words, systematically analysing the data by identifying and labelling meaningful items relevant to the research questions. The computer software NVivo was utilised in this process for more efficient coding. Table 2 shows a sample of codes generated after the initial round of data analysis.

Table 2. *An example of initial codes*

Interview transcripts and diaries	**Codes**
"Study by Topic allows me to speak on the themes. Its library is quite large, so I learned more words by topic. I think that increases the length of the sentences I say. This part also allows me to choose the topic freely. For example, if I'm tired, I'll choose something fun. If I'm not tired, I'll choose something difficult." (Participant 6)	Perceived useful features: Study by topic Practice with themes Learn topical vocabulary Freedom to choose topics
"The conversation practice was most useful to me because it simulated real talking situations. It helps me practice using different phrases in context, which in turn helps me speak more fluently and confidently. In addition, it also helps me practice taking long breaths to speak a long paragraph. I found the feedback very helpful. The feedback is very detailed and tells me exactly which sounds or words I need to improve. I appreciate feedback that gives me a score, so I can track my progress and understand where I need to focus." (Participant 1)	Perceived useful features: Conversation Simulate real conversations Practice using language in contexts More fluent More confident Practice speaking at length Detailed feedback Exact feedback Quantified feedback
"And I think Elsa's voice recognition is better, but when you're in an environment where it's just a little bit noisy, and when you need to study on the app, the results are affected quite a lot." (Participant 4)	Inaccuracies in recognition due to background noise

The analysis then advanced through multiple rounds of coding. After that, related codes were grouped and organised into themes that reflect meaningful patterns across the dataset in relation to the research questions (Braun & Clarke, 2006). Phases four and five of the thematic analysis focused on reviewing and refining

themes. This process was recursive as the researcher navigated back and forth between phases. Table 3 depicts how the themes addressing RQ2 were derived.

Table 3. *An example of emerging themes for RQ2*

Themes	**Sub-themes**	**Codes**
Perceived advantages	Increase trust and positivity towards AI-assisted learning	Can speak at length & with less hesitation
		Can speak spontaneously & with better intonation
		Improved communication skills
		Improved pronunciation - timely error correction
		Increased confidence
		Increased grammar accuracy
		Increased interest in practising speaking
		Increased vocabulary
		Personalised practice & feedback

Table 3. *(continued)*

Themes	**Sub-themes**	**Codes**
Perceived advantages	Enhance learning experience	Allow for repetition and multiple practice
		Help build learning habits
		Highlight strengths and weaknesses
		Integrate the functions of many apps into one
		Less pressure than speaking with humans
		Offer fun, enjoyable learning
	Provide an English speaking environment	Can learn how to ask questions
		Offer a systematic approach for learning
		Prepare for real-life interactions
		Various functions and topics for learning
	Offer detailed feedback	Detailed feedback with scores and percentages
		Display of the IPA recognition system for pronunciation adjustments
		More personalised feedback vs feedback in the classroom
		Suggestions for better phrases or vocab

Table 3. *(continued)*

Themes	**Sub-themes**	**Codes**
Perceived disadvantages	Learning-related challenges	Lack progress tracking - cannot review past lessons
		Limited practice reminders
		General feedback for fluency
		Strict/rigid/mechanical feedback
	Technical challenges	Delayed responses
		Fail to recognise voice
		Inaccuracies due to background noise
		Inconsistent user interface
		Using tricks to bypass the app's error spotting
	Content-related challenges	Avoid specialised topics - lack of knowledge
		Disconnection between the app and real life

As shown in Table 3, the two main themes for RQ2 were developed through the integration of multiple sub-themes, each comprising several related codes. King (2004) recommended that themes should not be considered final until all data have been thoroughly reviewed and all coding has been checked at least twice. In line with this recommendation, the coding process in the current study was conducted twice to ensure intra-coder reliability (Révész, 2023), and the themes were revisited multiple times to identify meaningful insights from the dataset.

The last stage of data analysis involved constructing a coherent narrative from all the themes to address the research questions, in which arguments were developed based on the researcher's interpretative analysis with data extracts integrated to support and illustrate key points.

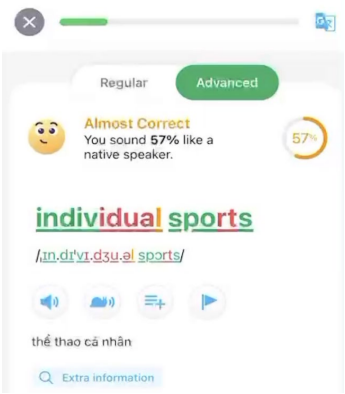
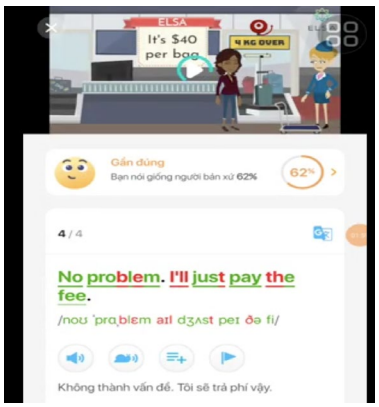
Data triangulation

Data triangulation refers to the use of multiple sources of data or methods of data collection to investigate the same research phenomenon from different perspectives (Flick, 2018). This method is suitable for this research because it enables a more holistic exploration of learner engagement with the ELSA application. For example, screen recordings captured participants' actual interactions with the app's interface, while diary entries offered reflective insights. Diary accounts (what participants reported) were cross-verified with screen recordings (what participants did) to ensure reliability and reduce potential biases. Table 4 illustrates how these data sources were triangulated to address Research Question 1.

Table 3. *An example of data triangulation*

Theme	**Screen recordings**	**Diary entries**	**Confirm/ Contradict**
Study by Topic	Participant 1 practised with topic-based lessons and long sentences. 	"I did daily practice exercises and then did pronunciation exercises and studied by topics. I find that the whole-sentence reading activity helped me speak more fluently. From there, I know where I should improve." (Participant 1)	Confirm

Table 4. *(continued)*

Theme	**Screen recordings**	**Diary entries**	**Confirm/ Contradict**
Improve Pronunciation	Participant 3 practised with the pronunciation feature for 15 minutes. 	"Today I practised Pronunciation and Intonation. I still want to improve my pronunciation because it is essential in communication." (Participant 3)	Confirm
Video Conversation	Participant 3 spent 20 minutes practising with a video conversation on a shopping topic. 	"Today I practised the Video Conversation feature. There is an additional video demonstration. I find that practising shadowing about studies, life, hobbies, etc., helps me feel less bored. From there, you can practice your reflexes and fluency in speaking English." (Participant 2)	Confirm

As shown in Table 4, 'Study by Topic' was identified in both the screen recording and the diary entry of Participant 1. The participant's engagement with topic-based lessons and sentence-reading activities, as observed in the recording, corroborated their self-reported learning activities, resulting in a "Confirm" classification. Conversely, if discrepancies were found between data sources, the classification would be marked as "Contradict." Similar patterns of confirmation were observed for *"Improve Pronunciation" (Participant 3) and "Video Conversation"* (Participant 2).

Trustworthiness

In qualitative research, trustworthiness is a concept introduced by Guba and Lincoln (1982) to replace notions of validity and reliability, which are rooted in positivist paradigms. Trustworthiness denotes the extent to which a study persuasively establishes the credibility of its findings. To ensure trustworthiness, the present study adhered to the following four critical criteria proposed by Lincoln and Guba (1985): Credibility, transferability, dependability and confirmability.

Credibility was ensured through multiple strategies. First, triangulating data sources—interviews, screen recordings, and diary entries—provided a clearer understanding of participants' perspectives. Prolonged engagement (i.e., a four-week data collection period) enabled the researcher to build rapport, better understand learners' behaviour, and minimise the possibility of data distortions arising from initial communications (Lincoln & Guba, 1985). Additionally, participants were allowed to write diary entries in Vietnamese to reduce linguistic barriers, thereby encouraging more authentic self-reports in the data.

Transferability was addressed through a thick description of the research setting, participant background, and the context in which the ELSA Speak app was used. This allows readers to judge the extent to which the findings might apply to similar contexts involving EFL learners and mobile-assisted language learning. Subsequently, *dependability* was enhanced through transparent documentation of the research procedures. Particularly, data analysis followed Braun and Clarke's (2022) thematic analysis, and coding was conducted and checked multiple times to ensure consistency and intra-coder reliability (Révész, 2023).

Finally, *confirmability* was established by connecting findings directly to the data through the inclusion of participant quotations and contextual descriptions. The use of multiple data sources enabled the researcher to construct themes from

participants' reports and actions, rather than imposing pre-existing assumptions. In line with Lincoln and Guba (1985), this approach assures readers that the research was conducted honestly and that the findings accurately reflect participants' voices and experiences.

FINDINGS

Research question (RQ) 1: What are perceived as functional features of ELSA Speak to improve English oral fluency by Vietnamese EFL university students?

The ELSA Speak application offers a range of features to support learners in developing various skills. Given that the current study focuses on English oral fluency, it is crucial to identify the ELSA Speak features that participants perceive as most beneficial for supporting this specific aspect of their language proficiency.

In the current study, participants used the ELSA Speak application over four weeks and recorded their experiences through device screen recordings, daily journal entries, and semi-structured interviews. Qualitative data analysis identified 13 ELSA Speak features that participants perceived as beneficial for improving oral fluency. These features are presented in Table 5, along with the frequency of mentions and the number of participants who referred to each.

Due to word limit constraints and the need for analytical depth, the answer to RQ1 focuses on the six most frequently referenced features: Study by Topic, Improve Pronunciation, ELSA AI Conversation, Practice Daily Lessons, Earn a Certificate – IELTS, and Conversation. However, it is important to note that excluding these features from this discussion does not imply their irrelevance. Participants meaningfully referenced all thirteen features, but the six presented here were mentioned more consistently, with some supported by more substantial data from both interviews and journal entries.

Table 6 summarises the key findings and ranks the six features in descending order of the number of times a feature is mentioned. Although the frequency of mentions may reflect the prevalence of these features in the study, it does not necessarily imply that one feature is more significant than others. Instead, the frequency may suggest a tendency for participants to perceive one feature as more or less relevant to them.

Table 5. Overview of ELSA Speak features identified as useful for improving oral fluency

Features	**Number of mentions**	**Number of participants who mentioned a feature**
Study by Topic	37	6
Improve Pronunciation	30	6
ELSA Al Conversation	25	5
Practice Daily Lessons	23	6
Earn a Certificate - IELTS	16	3
Conversation	14	3
Tests	3	2
Listening	3	2
Video Conversation	3	1
Study Sets	2	1
Speech Analysis	1	1
Unscrambling Words	1	1
Word Stress	1	1

Table 6. Perceived useful features of ELSA Speak for improving English oral fluency

Features	**Summary of participants' comments**	**Examples**
Study by Topic (37 mentions)	Participants found studying by topic particularly useful for developing their fluency through vocabulary improvement and practice with a range of themes.	"Study by Topic allows me to speak on the themes. Its library is quite large, so I learned more words by topic. I think that increases the length of the sentences I say. This part also allows me to choose the topic freely. For example, if I'm tired, I'll choose something fun. If I'm not tired, I'll choose something difficult." (Participant 6) "I think the activity that I found most useful to improve my fluency was studying by topic. One time, I studied online dating. I felt like I was practising speaking with a real person. When I practised speaking on this topic, I accumulated not only vocabulary but also knowledge." (Participant 8)

Table 6. (continued)

Features	**Summary of participants' comments**	**Examples**
Improve Pronunciation (30 mentions)	Participants repeatedly reported using the app to practice final sounds, stress, intonation, and specific phonemes. The app is useful for identifying and correcting their mistakes.	"I want to improve my pronunciation because it is very important in communication." (Participant 3) "Today I only had time for Daily Practice and Pronunciation Improvement (PI), which was useful when I found myself making a serious mistake." (Participant 6)
ELSA Al Conversation (25 mentions)	Participants highlighted the usefulness of interacting with AI, which provided natural and responsive practice environments. The AI feature was considered a helpful and engaging tool, offering role-play scenarios and also marking grammar and vocabulary for self-assessment.	"It trains me to think about what I need to say in a certain situation. And if there are some ideas that I don't know how to express, I will go back to ELSA AI Conversations and practice some words that I don't know." (Participant 3) "But in this conversation section, ELSA fixes long sentences and it counts fluency in its scoring as well. Other parts do not have criteria for fluency." (Participant 5)

Table 6. (continued)

Features	**Summary of participants' comments**	**Examples**
Practice Daily Lessons (23 mentions)	Participants reported that they frequently used Practice Daily , combined with ELSA AI and Study by Topic, to address specific weaknesses such as pronunciation or fluency.	"Today I did the Practice Daily . Reading whole sentences activity helped me speak more fluently. It causes me to catch my breath and focus more." (Participant 1) "I did Practice Daily because I wanted to focus on improving my pronunciation weaknesses, such as the /dʒ/ sound." (Participant 4)
Earn a certificate - IELTS (16 mentions)	Participants reported consistently using several lessons from the "Introduction to IELTS Band 6" and "Band 7" courses to enhance their speaking skills and achieve a higher test score.	"Today I practised with IELTS Speaking. I took the opportunity to try talking to AI to see if I could react quickly. Today I practised the topic "talking about a summer trip". The feedback was good; it told me what level I was at." (Participant 2) "I want to continue using this activity to improve my actual IELTS score." (Participant 4)

Table 6. (continued)

Features	**Summary of participants' comments**	**Examples**
Conversation (14 mentions)	Participants mentioned using Conversations to practice dialogue skills. There are specific topics, such as "Talk to your Colleagues" and "Talking to the Boss", where they can revisit vocabulary and useful phrases and integrate conversations into broader exercises involving pronunciation and stress.	"I found that practising dialogues was the most helpful for me because I practised topics related to everyday life. The dialogues were about sharing experiences about my loved ones, which helped me make connections and think faster while speaking." (Participant 5)

As shown in Table 6, the most frequently cited features were 'Study by Topic', followed by 'Improve Pronunciation' and 'ELSA AI Conversation.' Among these, 'Study by Topic' and 'ELSA AI Conversation' stood out in both journal and interview data and were observed to support fluency in different ways. These were followed by 'Practice Daily Lessons', 'Earn a Certificate – IELTS' and 'Conversation', which are explained in more detail below.

Study by Topic

Six out of seven participants appreciated 'Study by Topic' for its flexibility in choosing themes that align with their interests, for building vocabulary within specific topics, and for acquiring additional knowledge while simultaneously practising their English skills. As illustrated in Table 6, Participant 1 noted that ELSA Speak helped them "practice using different phrases in context," which contributed to their speaking "more fluently and confidently." Participant 8 found 'Study by Topic' to be the most useful activity for them, sharing that when practising a theme like online dating, they "felt like [they were] practising speaking with a real person" and were able to "accumulate not only vocabulary but also knowledge."

Notably, Participant 6 appreciated the feature's thematic structure and flexibility, explaining that its large topic library allowed them to "learn more words by topics," which they believed helped "increase the length of the sentences" they could produce. This result is meaningful because it aligns with the notion in previous research (e.g., Little, 2004; Ushioda, 2011) that a sense of control and empowerment in their own learning processes is likely to intrinsically motivate learners to persist in their learning.

Improve Pronunciation

Table 6 shows that the 'Improve Pronunciation' feature was valued for providing focused feedback on discrete phonological elements such as final sounds, stress, and intonation. Six participants reported using this to identify and correct pronunciation errors that impeded their fluency. Specifically, Participant 3 stated, "I want to improve my pronunciation because it is very important in communication." This view was echoed by Participant 6, who found this feature particularly useful for addressing serious errors during practice.

ELSA AI conversation

Although 'ELSA AI Conversation' was not mentioned as frequently as 'Study by Topic' or 'Improve Pronunciation' during the 4-week observation and diary entries, it was highlighted by five out of seven participants as significantly useful for improving their fluency during the post-interviews. This was largely attributed to its capacity to simulate realistic, interactive dialogues with an AI interlocutor, encouraging learners to think in the target language and practise unfamiliar words in real-life contexts.

As shown in Table 6, Participant 2 reported practising with the feature by "chatting with a virtual speaker" and appreciated that the AI "responded to [their] statements." Participant 8 described a similar experience, stating that "it was also like my friend, talking with me very naturally and quickly." For Participant 3, the feature encouraged spontaneous thinking and allowed for iterative practice: "It trains me to think about what I need to say in a certain situation," and when struggling with vocabulary, they would "go back to 'ELSA AI Conversations' and practice some words… then go back and practice again." This finding is significant because it confirms prior studies' findings on improving learners' spoken proficiency through fostering authentic and spontaneous conversations (Jung Youn, 2023; Tavakoli et al., 2020).

Participant 5 also emphasised that, unlike other features which focus primarily on word-level pronunciation, 'ELSA AI Conversation' provided feedback on longer utterances and incorporated "fluency in its scoring," which they noted was "not available in other parts." This outcome contributes to prior research on the usefulness of conversational AI in delivering such feedback, focusing on a broader spectrum of language proficiency dimensions rather than exclusively on pronunciation (Ji et al., 2022; Zechner & Hsieh, 2024), thereby demonstrably enhancing learner engagement and language learning outcomes.

Practice Daily Lessons

Reported by six of seven participants, the 'Practice Daily Lessons' feature was typically used in conjunction with others, particularly to build study habits and target specific weaknesses. Participant 1 reported that reading whole sentences helped enhance fluency, explaining that it required them to "catch [their] breath and focus more." Similarly, Participant 4 used the feature to target specific pronunciation challenges, such as the /dʒ/ sound. Participant 6 highlighted the motivating design of the activity, stating, "There is no excuse to stop grinding on Daily Practice," and described the exercises as "somewhat crucial." These insights suggest that Daily Practice supported both sustained engagement and targeted improvement in pronunciation.

On the one hand, the 'Improve Pronunciation' feature provided targeted feedback on specific phonological elements, helping learners identify and address individual errors. On the other hand, the 'Practice Daily Lessons' feature reinforced pronunciation through regular, structured activities, thereby helping learners build consistency and motivation in their study routines. On the other hand, these two features complement each other by supporting both accuracy and fluency in learners' speaking development. This suggests that participants viewed pronunciation accuracy as a key component of fluency development, aligning with previous research on teachers' and learners' perceptions of oral fluency (Khau & Huynh, 2022).

Earn a Certificate – IELTS

'Earn a Certificate – IELTS' was cited by three participants preparing for standardised speaking tests and was valued for its structured progression and scoring system, which allowed users to measure their performance levels and track improvements. Participant 2 used the feature to test their spontaneous speaking ability, explaining, "I took the opportunity to try talking to AI to see if I

could react quickly," and noted that the feedback was useful for identifying their speaking level. Participant 4 expressed a clear intention to continue using this activity, stating, "I want to continue using this activity to improve my actual IELTS score." These responses suggest that the feature not only supported speaking fluency but also served as a motivational tool for learners who aim to succeed in standardised tests such as IELTS.

Conversation

The final feature presented in Table 6 is 'Conversation', reported by three out of seven participants. Unlike 'ELSA AI Conversation', which allows free responses without models and provides feedback on overall performance, pronunciation, vocabulary, and grammar, the 'Conversation' feature offers sample dialogues for users to repeat, followed by immediate, percentage-based accuracy feedback. Participant 5 reflected that this feature helped them "think faster while speaking," while Participant 8 noted its usefulness as a preparatory step for more open-ended tasks, stating that when using 'ELSA AI Conversation', "I couldn't think of any ideas… I felt like I was stuck," and therefore preferred practising with structured dialogues first.

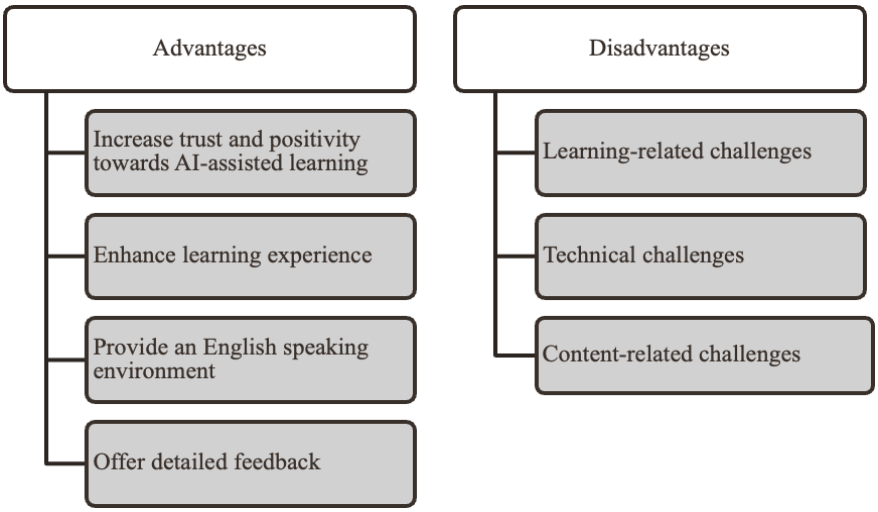
In summary, the current findings underscore the importance of integrating both activities that require learners to express their communicative intent in realistic interactional scenarios and tasks that focus on accuracy in AI-assisted language learning. This is consistent with prior research emphasising the role of simulated interaction, targeted pronunciation feedback, and personalised learning in developing oral fluency (e.g., Ebadi & Ebadijalal, 2022; Kormos & Dénes, 2004) as well as with the *Review of previous research* section.

Research question 2: What are Vietnamese EFL university students' general perceptions towards the advantages and disadvantages of using ELSA Speak for improving fluency?

While Research Question 1 focused on the specific features of ELSA Speak perceived as useful for fluency development, Research Question 2 expands the scope to explore participants' broader experiences using the app, thereby enabling the optimisation of its beneficial features while addressing its limitations. It is essential to note that the term "useful features" in RQ1 refers to components of the application (e.g., Study by Topic, Improve Pronunciation), whereas in RQ2, "advantages" and "disadvantages" refer to learners' experiences—the processes of engaging with ELSA Speak over time.

The findings for RQ2 presented in this section are based on a content analysis of participants' interview transcripts and journal entries, triangulated with screen recordings of participants' practice over the four-week study. Thematically coded, the combined data reveal patterns highlighting both the app's benefits and limitations. Figure 5 presents a summary of these findings.

Figure 5. Advantages and disadvantages of using ELSA Speak for developing English oral fluency as perceived by the participants



Advantages of ELSA Speak

Figure 5 illustrates four key advantages that emerged from the analysis: increased trust and positivity toward AI-assisted learning, enhanced learning experience, provision of an English-speaking environment, and detailed feedback. In the subsequent paragraphs, each of these themes will be discussed in more detail, accompanied by notable excerpts from the participants.

Increase trust and positivity towards AI-assisted learning

First of all, all participants reported that using ELSA Speak has increased their trust in the tool and helped them have a more favourable perception of employing AI-supported technology for learning purposes, as Participant 1 asserted, "My trust in the app has increased. I think I am more positive about practising with an app than before."

This phenomenon may be attributed to the participants' experiences of not only improvements in grammatical and lexical accuracy, pronunciation, and fluency, but also an increase in their interest and confidence in speaking, especially in articulating extended sentences. Along this line, Participant 4 shared:

> I have seen some improvements. In ELSA AI, I feel like I'm talking to a real person. That's also a way for me to be more confident when I practice. I feel less nervous when I'm talking to people. I can see this very clearly in my interactions in English class at school.

It is noted that following a period of practice with the ELSA Speak application, Participant 4 experienced a noticeable increase in confidence and a decrease in nervousness in real-life interactions. This finding is significant because it reveals an app's impact beyond controlled practice sessions, extending to real-life communication skills. Besides that, it corroborates prior research on the relationship between AI-assisted language learning and students' affective factors, encompassing constructs such as confidence and motivation (Wang et al., 2024).

Enhance learning experience

As seen from Figure 5, all participants appreciated the ELSA Speak app's ability to enhance their learning experience. This is because it provides them with a range of functions and topics, presented in a systematic, structured manner. On this point, Participant 5 reported, "There are so many lessons in the app, and it covers a whole range of speaking skills. Every time I come in, I will have a different practice." Participant 2 added, "ELSA suggests a systematic approach, like the IELTS part has lessons for band 6.0 to band 9.0. It is very clear." This remark underscores the ability of the application to deliver a scaffolded learning experience by systematically sequencing the lessons with progressively increasing difficult levels.

The findings are consistent with existing research exploring the role of AI-powered tools in language education. Such tools demonstrate the capacity to personalise language learning by providing quantitative assessments of speaking performance, while also assessing learners' weaknesses, strengths and knowledge gaps to suggest practices that meet their needs (e.g., Ebadi & Ebadijalal, 2022; Wang et al., 2023).

Provide an English speaking environment

The analysis also revealed that seven participants agreed that ELSA Speak provided an environment for them to practice speaking English, which they often found lacking in both their classrooms and social settings in Vietnam. Participant 8 revisited their experience:

> [...] When I was in high school, I just studied grammar, reading, writing, etc. I didn't practice speaking. I didn't have a good environment. Now, when I started practising with ELSA, I realised I was speaking every day. It not only formed a habit for me but also helped me correct my pronunciation and speak more fluently. I used to stumble a lot, but now I can speak more fluently.

This indicates that, through the platform, Participant 8 benefited from an environment that allowed them to practice speaking daily, which contributed to the development of their habit, improved pronunciation, and enhanced overall fluency, with reduced hesitation. This insight is noteworthy because it echoes earlier research by de Jong and Perfetti (2011), who stressed that repeated practice is necessary for proceduralisation, making speaking automatic and improving fluency.

Offer more detailed feedback

Finally, six participants indicated that ELSA Speak offered more detailed and comprehensive feedback, particularly compared to human teachers or other applications. Participant 2 recounted, "ELSA AI helps me remember the old lessons and it also gives me good feedback. It's much more detailed than the ChatGPT app on my mobile phone." Participant 8 asserted their experiences as follows:

> The feedback for me was wonderful. When I come across a sentence I don't know how to say, ELSA will provide me with a sample sentence that includes pronunciation and stress guidance. After I finish speaking, it will give me a percentage indicating how well I performed. This is something I rarely encountered when I was studying in school and receiving too general comments from teachers or friends.

It can be observed that the participant benefited from a distinct advantage of ELSA Speak: providing learners with sample models and quantitative performance metrics, such as percentage scores. This contrasts with traditional classroom

settings in countries such as Vietnam, where teacher or peer feedback may be less specific. This was again stressed by Participant 2:

> ELSA Speak's feedback is detailed, clearly outlining percentages and guidance on how to improve. Compared to the reality of going to school, teachers often focus on our listening and understanding of what they say, and the class is quite crowded, so we don't get much feedback on fluency and pronunciation. Besides, I don't often practise speaking in class, and when I do, it's just a little, so real-life feedback is insignificant. That's why studying at home with Elsa Speak helps me gain a deeper appreciation.

Disadvantages of ELSA Speak

Despite the advantages presented above, the data analysis also identified several limitations of the ELSA Speak application. As seen in Figure 5, three prominent themes include learning-related, technical and content-related challenges. These themes will be presented in greater detail in the following sections, with direct quotes from participants.

Learning-related challenges

As shown in Figure 5, challenges with using ELSA Speak for learning (e.g., inability to track past learning history, lack of reminders, overly general or overly rigid feedback) were reported by 6 participants. Among these, three participants highlighted the lack of records for completed lessons, making it difficult to review past exercises or track improvements. Regarding this, Participant 3 shared, "For the disadvantage, I cannot find the history of what I learned the previous day. It's hard to follow up the next day." Similarly, Participant 2 asserted that:

> ELSA AI doesn't store the past lessons. Sometimes, I want to go back and see if I have improved, or if I've used a lot of words I've just learnt; there are no records. If I forget to take a picture, I won't be able to see it again. Also, many topics for ELSA AI Conversations change daily and are not fixed. So, if I revisit the app another day, that topic will no longer be there. I can't go back and practice that topic again.

What has been observed from these participants' recounts is that ELSA Speak appears limited in its capacity to support long-term learning. Indeed, its lack of records of learning history has been demonstrated to have inhibited learners from

revisiting and reviewing previous lessons and from practising certain content and reinforcing their skills.

This finding may contribute to existing research by addressing longer-term use of an application. This is unlike studies with only brief interventions - e.g., one hour with Google Assistant (Chen et al., 2023) or six hours with speech recognition websites (Bashori et al., 2024) - which may be biased due to novelty effects, where initial excitement with a technology can lead to better outcomes than those produced by sustained use (van den Berghe et al., 2019).

Technical challenges

The second challenge, as depicted in Figure 5, concerns the technical functionality of ELSA Speak, as highlighted by four participants. In particular, a noticeable delay was reported in registering user input. Concerning this, Participant 6 pointed out:

> Sometimes the buttons lag. For example, sometimes I complete multiple-choice questions (MCQs), and the app does not receive the answer I choose until 3 seconds later. It was a bit weird. Three seconds felt so long.

Being a student specialising in computer science, Participant 6 provided further insight into the cause of this delay:

> The bad thing is that the user interface is not synchronised. For example, in this lesson, the interface is A. In the other lesson, the interface is B. So, when you use it, the experience is not very good. When I first opened the app, I had to search for a long time before I found the ELSA AI function. I think the reason is that the app doesn't have a guide for its interface. It doesn't tell you what each button does.

Based on Participant 6's observations, the application's design remains limited, with noticeable technical delays, inconsistent interfaces, and insufficient user guidance, which may eventually undermine user satisfaction and experience. This aligns with findings reported in prior studies, such as Jeon (2022).

Another critical issue is the accuracy of ELSA Speak's speech recognition function. Relatedly, Participant 8 remarked, "Although I have practised the sound, sometimes four times, five times, it still marks my pronunciation wrong. In such a case, I will move on and not practise anymore." Similarly, Participant 2 reported,

"When you're in an environment where it's a little bit noisy, and when you need to study on the app, the results are affected quite a lot."

Along the same line, but more strikingly, Participant 1 underscored a major problem with how the app assesses pronunciation, yet prioritises volume over correct syllable stress, which led to their diminishing interest in continuing to use the app:

> The biggest disadvantage is that when it comes to stressing the syllables, you don't have to stress them properly. Instead, you only need to read them really loud and it will let you pass. It just lets go of the mistakes. When I discovered that loophole, I lost interest. I don't want to do similar exercises anymore.

The insights from the participants revealed that, consistent with previous research (e.g., Zou et al., 2023), ELSA Speak struggles with reduced accuracy under less-than-ideal conditions, such as noise, and occasionally fails to accurately recognise pronunciation despite multiple attempts. Notably, these results demonstrate that participants abandoned practice sessions due to frustration.

Content-related challenges

Although only two of seven participants highlighted challenges with the ELSA Speak application's capacity to effectively engage with the content of their conversations, their observations are worth exploring due to their potential impact on user experience, given the app's limitations. For instance, while Participant 6 expressed a desire to engage in in-depth discussions about computers, ELSA AI fell short of their expectations, struggling to handle technical or specialised topics:

> I tried to explain, but ELSA didn't understand. The way ELSA AI did is to avoid and divert into another topic instead. It's quite limited if we go deeper into technical knowledge. Another thing is that that when I need to practice negotiating for a work contract, ELSA would probably give up. It can only talk about common knowledge.

This constraint may be attributed, as Amaral and Meurers (2011) concluded, to the lack of contributions from language educators during the development phase of an application or system. Consequently, to ensure learners sustain their ongoing intention to utilise the ELSA Speak application, it is important that these shortcomings be addressed and mitigated.

DISCUSSION

Over the course of a four-week study, data collected through observations, screen recordings, daily journal entries, and interviews with participants revealed that students do not passively use language learning applications. Rather, they deliberately and actively select features that meet their needs—in this case, the development of oral fluency. The ELSA Speak features that learners found particularly useful were those that offered them autonomy or flexibility to make choices, addressed their personal fluency goals, and provided meaningful, detailed feedback that extended beyond the pronunciation accuracy of individual words. For example, among many features, 'Study by Topic' was valued not only for vocabulary input but also for allowing flexible topic choice, while 'ELSA AI Conversation' supported spontaneous thinking and fluency-building through context-based conversations and the assessment of fluency performance.

The above insight highlights that, firstly, in an AI-mediated learning environment, affective factors such as feeling confident and in control remain crucial to sustaining learner engagement. This aligns with the arguments of Little (2004) and Ushioda (2011) that a sense of control and empowerment in one's own learning processes is likely to intrinsically motivate learners to persist in their learning. Secondly, the finding supports Tavakoli et al. (2020) and Jung Youn (2023), who underscored the importance of authentic tasks in developing learners' spoken proficiency and Xie et al. (2019) and Chiu et al. (2024), who demonstrated that multidimensional and adaptive personalised feedback promotes greater learner engagement and outcomes.

The study also found that participants benefited from strategically combining various ELSA Speak features in a coordinated manner. For instance, they were observed regularly using Practice Daily Lessons to support the improvements targeted by 'Improve Pronunciation'. On the one hand, this suggests that learners perceive fluency as intertwined with or interdependent on pronunciation (Khau & Huynh, 2022). On the other hand, feature integration has frequently been observed in language learning applications and has been shown to support multiple dimensions of linguistic proficiency rather than focusing on a single aspect. Such integration has been shown to lower anxiety, promote learners' engagement and increase their willingness to communicate (Hatmanto & Sari, 2023; Zhao et al., 2024).

The findings from the current study indicate that Vietnamese EFL university students generally hold positive perceptions toward ELSA Speak as a tool for improving English fluency, particularly due to its ability to enhance confidence, provide structured learning, offer consistent practice opportunities, and deliver detailed feedback — areas often found lacking in Vietnamese classroom settings (Tran & Tanemura, 2020).

Wang et al. (2024) found that learners who felt more engaged with AI tools and viewed them positively experienced greater learning enjoyment and achieved better outcomes. Additionally, Shafiee Rad (2024) and Mei et al. (2024) found a positive relationship between AI-supported tools and affective variables, such as confidence and motivation. The current study supports these findings in several ways. Firstly, differing attitudes between Participant 6 (more positive) and Participant 5 (less positive) appeared to influence the quality of their experiences with the app, reinforcing the role of affective engagement in learning outcomes. Secondly, participants reported increased self-efficacy and reduced nervousness in their real-life interactions, suggesting that ELSA Speak's impact extends beyond controlled practice and affects users' real-life communication skills.

Research has highlighted the capacity of AI-powered language tools to deliver scaffolded learning experiences, in which their content is organised to support learners' gradual skill development (Hatmanto & Sari, 2023). Findings showed that ELSA Speak reflects this principle through its diverse lessons and topics, presented systematically and in a structured way. Along with more detailed and comprehensive feedback, especially compared to that in traditional classrooms, participants reported benefiting from knowing their strengths and weaknesses, and consequently improving by following personalised practice suggestions that met their needs (Ebadi & Ebadijalal, 2022).

Despite its strengths, ELSA Speak has its limitations. A notable drawback is that it may not be suitable for long-term learning because it lacks a learning history feature. Reported by six participants, this limitation prevented them from revisiting past lessons or reinforcing specific skills, which may ultimately hinder the effectiveness of spaced repetition and long-term retention. This concern is relevant because it has not been adequately addressed in short-term studies (e.g., Bashori et al., 2024; Chen et al., 2023), which may thereby underestimate the challenges related to long-term engagement (van den Berghe et al., 2019).

Aligned with earlier criticisms of AI-based learning tools such as Zou et al. (2023) and Jeon (2022), participants in the current study also raised issues with ELSA Speak's technical performance, such as slow loading times, confusing interface, lack of specialised knowledge and most importantly, inconsistent recognition, which led to their frustration and abandonment of practice sessions. These issues are especially critical for users in under-resourced or noisy environments.

Notably, the notion that loud speech, rather than accuracy, could trick the system into awarding a high score (highlighted in the "Disadvantages of ELSA Speak" section) raises ethical and pedagogical questions about the reliability of AI feedback. The current study, therefore, aligns with previous research that the integration of AI-powered tools in language education must be pedagogically grounded and critically mediated, rather than implemented uncritically or in isolation (Yoshida, 2018).

Contributions of the current study

The current study has contributed to theoretical, methodological, and practical discussions in the field of Mobile-Assisted Language Learning (MALL) in TESOL, particularly in the context of AI-Assisted Language Learning (AILL).

Theoretical contributions

At the theoretical level, the current study makes several contributions to understanding MALL. Firstly, it affirms the usefulness of features that promote learner autonomy, authentic communication and comprehensive feedback for fluency development. Secondly, rather than viewing fluency as limited to speech rate or the number of pauses, as conceptualised by Arevart and Nation (1991) and Schmidt (1992), the current study highlights how learners perceived it to relate to motivation, confidence, and their strategic use of technological resources. This perspective adds depth to fluency research by emphasising the roles of affective (Dörnyei & Ryan, 2015) and metacognitive factors (Qiao & Zhao, 2023) in shaping learners' engagement with AI-driven language tools for fluency improvement.

Thirdly, the study sheds light on the challenges associated with the relatively long-term usability of ELSA Speak, raising concerns about novelty effects in prior research and suggesting the need for longitudinal studies to observe lasting effects (van den Berghe et al., 2019). Finally, it adds valuable cultural and contextual insight to the MALL literature by focusing on a Vietnamese university context—a relatively underrepresented population in MALL research (Ngoc & Dung, 2020). By

doing so, it expands the geographic and cultural scope of existing research, offering diverse perspectives on how learners from Southeast Asia interact with educational technologies in resource-constrained yet digitally connected environments.

Methodological contributions

The current study makes methodological contributions to MALL research in several ways. It demonstrates the value of employing a qualitative design that combines several data collection tools such as screen recordings, reflective journals, and semi-structured interviews. This approach enabled the capture of a rich understanding of how Vietnamese EFL university students engaged with the ELSA Speak app over an extended period. For example, while screen recordings revealed patterns of feature use and navigation behaviour, journal entries documented participants' daily reflections on progress, challenges, and motivations. According to Sántha and Malomsoki-Sántha (2023), this triangulation of data sources offered a more holistic view of the observed phenomenon than single-method studies.

Furthermore, the instruments developed for the current study, especially the journal prompts and interview questions, could elicit detailed perceptions of app features, usage strategies, motivation, progress, advantages, and disadvantages, among others. Researchers interested in exploring learners' interactions with mobile technologies may find the approach and instruments employed herein replicable and adaptable to other contexts or populations, especially when in-depth insights into learners' experiences are needed. Future researchers may also adapt the study's themes and findings to design questionnaires for quantitative or mixed methods studies, especially those aiming to quantify learners' preferences or engagement patterns in AILL contexts. In this way, the study contributes to the early stages of instrument development, thereby supporting the expansion of MALL research.

Notably, the study offers insights into managing participant engagement in future research. One challenge during the data collection phase was maintaining learner participation over a four-week period. In the first week, participants engaged enthusiastically, but from the second week onward, they began skipping practice sessions occasionally. Following the strategies suggested by Killien and Newton (1990) to address this, the researcher established a strong connection with participants from the start, showed appreciation for their time, and explained how

their involvement would contribute to the study's goals. Additionally, the participants were clearly informed about the study's duration, time commitment, and what participation entails, to mentally prepare and set realistic expectations before giving consent to participate. As such, with clear instructions, regular check-ins and gentle weekly or daily reminders, their engagement was quickly reestablished.

Finally, the study contributes to ongoing discussions about methodological trustworthiness in qualitative research (Holliday, 2015). By systematically aligning data sources and cross-verifying findings across different modes, the current study substantiates how triangulation can enhance the credibility of learner-centred research.

Pedagogical contributions

The findings of the current study offer several implications for English language educators. Specifically, the results highlight the potential for integrating ELSA Speak into classroom instruction by integrating specific ELSA Speak features into in-class learning objectives. For example, the 'Study by Topic' feature could be aligned with weekly speaking themes such as travel, technology, or relationships. Teachers could assign these topics for pre-class preparation and then apply the vocabulary and expressions practised through the app in in-class discussions, roleplays, or debates. This helps to address the challenge of limited in-class practice time often encountered in traditional crowded classroom environments.

As shown in the current study, participants often combined several features in a single use; for example, 'Daily Practice Lessons' with 'Improve Pronunciation'. However, some students may require initial guidance with navigating different features, interpreting app feedback or creating a study plan based on individual fluency goals. Educators could, for instance, conduct training workshops at the beginning of the term to guide students in using ELSA Speak more effectively. This helps to avoid frustration or disengagement due to misunderstandings about how the app functions, consistent with Hoi and Mu (2021).

To mitigate app limitations raised by participants, educators can play a mediating role by shifting their mindset, helping them move away from viewing ELSA Speak as a scoring system and instead embracing it as a resource for learning. In this case, learners reported that reading loudly could occasionally mislead the app into assigning higher pronunciation scores. Teachers could encourage them to treat ELSA Speak as a feedback tool and prompt them to pay close attention to, for

instance, the example audio, analyse it carefully, think about how their own pronunciation or speech compares to it, and focus on real progress rather than just marks. This can then be reinforced through peer-review activities, where students listen to each other's recordings and discuss pronunciation, stress, and fluency issues.

Finally, the reflective practices used in the current study—particularly the guided journal prompts—can be adapted for classroom use. Teachers may encourage students to keep a short learning journal (either digitally or in print) in which they reflect on the app features they used, why they chose them, and what they learned, thereby fostering learners' critical thinking and metacognitive awareness (Kessler, 2023). These reflections can then be discussed weekly in small groups or submitted for formative feedback, promoting learner autonomy and self-regulation (Qiao & Zhao, 2023).

Limitations and future research

While the current study offers valuable insights into learner engagement with ELSA Speak, several limitations must be acknowledged. The first limitation concerns its size. With a small sample size (n = 7), the findings are not intended to be generalisable to the wider population of EFL learners. To address this, future research could involve larger, more diverse participant groups, spanning different regions and age ranges, to capture more comprehensive learner experiences and perceptions. Another limitation is the relatively short study duration (4 weeks), which may not be sufficient to observe long-term changes in learners' behaviour or the sustained impact of ELSA Speak. Future studies could adopt a longitudinal design to enable researchers to track changes in fluency development and app engagement over a longer period.

Additionally, the study focused on a specific context — Vietnamese EFL university students — which may limit the transferability of its findings to learners in other contexts. Further research should conduct similar studies in different regions to determine whether learners in other contexts perceive and use ELSA Speak in similar ways. Finally, further research might investigate the role of teachers in integrating ELSA Speak into blended learning environments. It would be valuable to examine how educators perceive the app, how they support learners' use of it, and what challenges they face in incorporating such tools into their curriculum. This could inform the development of technological designs better aligned with educational needs and real-world learning and teaching contexts.

CONCLUDING REMARKS

The present study has empirically substantiated the growing body of research on AILL by providing an in-depth exploration of Vietnamese EFL university students' perceptions and experiences with the ELSA Speak application. It has underscored the application's potential to develop oral fluency through authentic practice, personalised feedback and motivational design, among others, while also revealing key strengths and limitations based on learners' firsthand engagement with the tool. In many ways, ELSA Speak functions like a language trainer that never sleeps, always available to offer feedback, repetition, and encouragement. However, as with any trainer, learners' success hinges on multiple factors, most notably their motivation and persistence. To borrow an analogy, AI-mediated apps are like GPS systems, since they can guide, suggest, and reroute, but the learner must be the one who drives. Over-dependence on GPS can lead to a loss of critical thinking about direction.

In the age where AI is increasingly involved in education, the current study serves as a timely reminder: technology can support, but not replace, the human aspects of language learning. While ELSA Speak can scaffold fluency, the findings revealed several significant limitations, reinforcing the perspective that it cannot fully substitute the role of human teachers in language instruction. Looking ahead, as AI technologies continue to advance rapidly, so must our frameworks for understanding and integrating them into practice. Learners, particularly those who are younger or less experienced, will need the right support and guidance. Ultimately, the idea is not simply to use an AI application, but to do so in ways that empower learners and preserve their original and critical thinking.

THE AUTHOR

Diem Thi Ngoc Tran is an English language educator and researcher based in Vietnam. She recently completed her Master of Education (TESOL) at the University of Sydney. She currently teaches English at a language centre in Ho Chi Minh City, where she designs innovative and interactive materials for learners of all levels.

ACKNOWLEDGEMENT

The author wishes to express sincere gratitude to Professor Aek Phakiti for his patient guidance and insightful feedback throughout the research process. Special thanks are extended to the participants for their valuable time and contributions, which made this study possible. The author is also deeply thankful to family and friends for their unwavering encouragement and support during the completion of this work. Heartfelt thanks are given to God for His sustaining grace and to the Christian community in Sydney and Ho Chi Minh City for their prayers and fellowship throughout this academic journey.

REFERENCES

Ahmed, A. A. A., Hassan, I., Pallathadka, H., Keezhatta, M. S., Noorman Haryadi, R., Al Mashhadani, Z. I., Yahya Attwan, L., Rohi, A., & Namaziandost, E. (2022). MALL and EFL learners' speaking: Impacts of Duolingo and WhatsApp applications on speaking accuracy and fluency. *Education Research International, 2022*, Article 6716474. https://doi.org/10.1155/2022/6716474

Amaral, L. A., & Meurers, D. (2011). On using intelligent computer-assisted language learning in real-life foreign language teaching and learning. *ReCALL, 23*(1), 4–24. https://doi.org/10.1017/S0958344010000261

An, X., Chai, C. S., Li, Y., Zhou, Y., & Yang, B. (2023). Modeling students' perceptions of artificial intelligence assisted language learning. *Computer Assisted Language Learning*, 1–22. https://doi.org/10.1080/09588221.2023.2246519

Arevart, S., & Nation, P. (1991). Fluency improvement in a second language. *RELC Journal, 22*(1), 84–94.

Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2024). "Look, I can speak correctly": Learning vocabulary and pronunciation through websites equipped with automatic speech recognition technology. *Computer Assisted Language Learning, 37*(5–6), 1335–1363. https://doi.org/10.1080/09588221.2022.2080230

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*(1), 579–616. https://doi.org/10.1146/annurev.psych.54.101601.145030

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. SAGE Publications Ltd.

Carson, J. G., & Longhini, A. (2002). Focusing on learning styles and strategies: A diary study in an immersion setting. *Language Learning, 52*(2), 401–438. https://doi.org/10.1111/0023-8333.00188

Chen, H. H.-J., Yang, C. T.-Y., & Lai, K. K.-W. (2023). Investigating college EFL learners' perceptions toward the use of Google Assistant for foreign language learning. *Interactive Learning Environments, 31*(3), 1335–1350. https://doi.org/10.1080/10494820.2020.1833043

Chiu, T. K. F., Moorhouse, B. L., Chai, C. S., & Ismailov, M. (2024). Teacher support and student motivation to learn with artificial intelligence (AI) based chatbot. *Interactive Learning Environments, 32*(7), 3240–3256. https://doi.org/10.1080/10494820.2023.2172044

de Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning, 61*(2), 533–568. https://doi.org/10.1111/j.1467-9922.2010.00620.x

Dizon, G., Tang, D., & Yamamoto, Y. (2022). A case study of using Alexa for out-of-class, self-directed Japanese language learning. *Computers and Education: Artificial Intelligence, 3*, 100088. https://doi.org/10.1016/j.caeai.2022.100088

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.

Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited.* Routledge.

Duff, P. A. (2010). Research approaches in applied linguistics. In R. B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (2nd ed.). Oxford University Press.

Ebadi, S., & Ebadijalal, M. (2022). The effect of Google Expeditions virtual reality on EFL learners' willingness to communicate and oral proficiency. *Computer Assisted Language Learning, 35*(8), 1975–2000. https://doi.org/10.1080/09588221.2020.1854311

ELSA Speak. (n.d.). *Product*. ELSA Speak. https://elsaspeak.com/en/product/

Evers, K., & Chen, S. (2020). Effects of an automatic speech recognition system with peer feedback on pronunciation instruction for adults. *Computer Assisted Language Learning, 35*(8), 1869–1889. https://doi.org/10.1080/09588221.2020.1839504

Ezzy, D. (2010). Qualitative interviewing as an embodied emotional performance. *Qualitative Inquiry, 16*(3), 163–170. https://doi.org/10.1177/1077800409351970

Farrokhi, F., & Mahmoudi, A. (2014). A socio-cognitive approach to developing oral fluency and naturalness in Iranian EFL learners. *International Journal of Applied Linguistics & English Literature, 3*(2), 1–15. https://doi.org/10.7575/aiac.ijalel.v.3n.2p.1

Fathi, J., & Rahimi, M. (2024). Utilising artificial intelligence-enhanced writing mediation to develop academic writing skills in EFL learners: A qualitative study. *Computer Assisted Language Learning*, 1–40. https://doi.org/10.1080/09588221.2024.2374772

Flick, U. (2018). Triangulation in data collection. In U. Flick (Ed.), *The SAGE handbook of qualitative data collection* (pp. 527–544). SAGE Publications Ltd. https://doi.org/10.4135/9781526416070.n34

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2019). *How to design and evaluate research in education* (10th ed.). McGraw-Hill Education.

Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second-language learning*. Newbury House Publishers.

Grimshaw, J., & Cardoso, W. (2018). Activate space rats! Fluency development in a mobile game-assisted environment. *Language Learning & Technology, 22*(3), 159–175. https://doi.org/10.10125/44662

Guba, E. G., & Lincoln, Y. S. (1982). Epistemological and methodological bases of naturalistic inquiry. *Educational Communication & Technology, 30*(4), 233–252. https://doi.org/10.1007/bf02765185

Hatmanto, E. D., & Sari, M. I. (2023). Aligning theory and practice: Leveraging Chat GPT for effective English language teaching and learning. *E3S Web of Conferences, 440*, 5001. https://doi.org/10.1051/e3sconf/202344005001

Heycke, T., & Spitzer, L. (2019). Screen recordings as a tool to document computer-assisted data collection procedures. *Psychologica Belgica, 59*(1), 269–280. https://doi.org/10.5334/pb.490

Ho, W. Y. J. (2021). 'I knew that you were there, so I was talking to you': The use of screen-recording videos in online language learning research. *Qualitative Research, 21*(1), 120–139. https://doi.org/10.1177/1468794119885044

Hoi, V. N., & Mu, G. M. (2021). Perceived teacher support and students' acceptance of mobile-assisted language learning: Evidence from Vietnamese higher education context. *British Journal of Educational Technology, 52*(2), 879–898. https://doi.org/10.1111/bjet.13044

Holliday, A. (2015). Qualitative research and analysis. In B. Paltridge & A. Phakiti (Eds.), *Research methods in applied linguistics: A practical resource* (pp. 67–81). Bloomsbury.

Istrate, A.-M. (2018). Artificial intelligence and machine learning – Future trends in teaching ESL and ESP. *eLearning and Software for Education, 14*(2), 471–476.

Jeon, J. (2022). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning, 37*(1–2), 1–26. https://doi.org/10.1080/09588221.2021.2021241

Ji, H., Han, I., & Ko, Y. (2022). A systematic review of conversational AI in language education: Focusing on the collaboration with human teachers. *Journal of Research on Technology in Education, 55*(1), 48–63. https://doi.org/10.1080/15391523.2022.2142873

Jung Youn, S. (2023). Test design and validity evidence of interactive speaking assessment in the era of emerging technologies. *Language Testing, 40*(1), 54–60. https://doi.org/10.1177/02655322221126606

Kaiser, D. (2024). Artificial intelligence integration in three iOS pronunciation apps: ELSA Speak, Loora, and Vocal Image. *Journal of Second Language Pronunciation, 10*(3), 404–426. https://doi.org/10.1075/jslp.24052.kai

Karim, S. A., Hamzah, A., Anjani, N. M., Prianti, J., & Sihole, I. G. (2023). Promoting EFL students' speaking performance through ELSA Speak: An artificial intelligence in English language learning. *Journal of Languages and Language Teaching, 11*(4), 655. https://doi.org/10.33394/jollt.v11i4.8958

Kessler, M. (2023). Supplementing mobile-assisted language learning with reflective journal writing: A case study of Duolingo users' metacognitive awareness. *Computer Assisted Language Learning, 36*, 1040–1063. https://doi.org/10.1080/09588221.2021.1968914

Khau, A. H., & Huynh, V. T. M. (2022). An investigation into oral fluency perceived by teachers and students—in a Vietnamese context of English education. *Language Testing in Asia, 12*(1), 1–19. https://doi.org/10.1186/s40468-022-00174-5

Killien, M., & Newton, K. (1990). Longitudinal research—The challenge of maintaining continued involvement of participants. *Western Journal of Nursing Research, 12*(5), 689–692. https://doi.org/10.1177/019394599001200512

King, N. (2004). Using templates in the thematic analysis of text. In C. Cassell & G. Symon (Eds.), *Essential guide to qualitative methods in organizational research* (pp. 257–270). Sage Publications.

Kirkpatrick, A. (2010). *English as a lingua franca in ASEAN: A multilingual model*. Hong Kong University Press.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32*(2), 145–164. https://doi.org/10.1016/j.system.2004.01.001

Kuddus, K. (2022). Artificial intelligence in language learning: Practices and prospects. In A. K. Tyagi, A. Mire, & S. Malik (Eds.), *Advanced analytics and deep learning models* (pp. 3–17). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119792437.ch1

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage Publications.

Little, D. (2004). Democracy, discourse, and learner autonomy in the foreign language classroom. *Utbildning Och Demokrati, 13*(3), 105–126. https://doi.org/10.48059/uod.v13i3.784

Ma, R., & Oxford, R. L. (2014). A diary study focusing on listening and speaking: The evolving interaction of learning styles and learning strategies in a motivated, advanced ESL learner. *System, 43*(1), 101–113. https://doi.org/10.1016/j.system.2013.12.010

McCrocklin, S. (2019). Learners' feedback regarding ASR-based dictation practice for pronunciation learning. *CALICO Journal, 36*(2), 119–137. https://doi.org/10.1558/cj.34738

Mei, B., Qi, W., Huang, X., & Huang, S. (2024). Speeko: An artificial intelligence-assisted personal public speaking coach. *RELC Journal, 55*(2), 596–600. https://doi.org/10.1177/00336882221107955

Melhuish, K., & Falloon, G. (2010). Looking to the future: M-learning with the iPad. *Computers in New Zealand Schools: Learning, Leading, Technology, 22*(3), 1–16.

Ngoc, D., & Dung, T. (2020). Key factors influencing learners' oral fluency in English speaking classes: A case at a public university in Viet Nam. *VNU Journal of Foreign Studies, 36*(6). https://doi.org/10.25073/2525-2445/vnufs.4631

Porto, M. (2007). Learning diaries in the English as a foreign language classroom: A tool for accessing learners' perceptions of lessons and developing learner autonomy and reflection. *Foreign Language Annals, 40*(4), 672–696. https://doi.org/10.1111/j.1944-9720.2007.tb02887.x

Qiao, H., & Zhao, A. (2023). Artificial intelligence-based language learning: Illuminating the impact on speaking skills and self-regulation in Chinese EFL

context. *Frontiers in Psychology, 14*, Article 1255594. https://doi.org/10.3389/fpsyg.2023.1255594

Rad, H. S. (2024). Revolutionizing L2 speaking proficiency, willingness to communicate, and perceptions through artificial intelligence: A case of Speeko application. *Innovation in Language Learning and Teaching*, 1–16. https://doi.org/10.1080/17501229.2024.2309539

Révész, A. (2023). Coding second language data validly and reliably. In A. Révész, A. Mackey, & S. M. Gass (Eds.), *Current approaches in second language acquisition research* (pp. 257–275). John Wiley & Sons, Inc. https://doi.org/10.1002/9781394259670.ch12

Rose, H. (2020). Diaries and journals: Collecting insider perspectives in second language research. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 386–396). Routledge.

Rudnik, Y. (2024). The use of artificial intelligence chatbots in teaching foreign languages as an innovative interactive technology. *Osvitolohichnyĭ Dyskurs, 45*(2). https://doi.org/10.28925/2312-5829.2024.2.2

Sántha, K., & Malomsoki-Sántha, Á. (2023). Methodological aspects for applied linguistic research – Mixed methods, triangulation. *Acta Universitatis Sapientiae, Philologica, 15*(2), 1–13. https://doi.org/10.2478/ausp-2023-0013

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition, 14*(4), 357–385. https://doi.org/10.1017/S0272263100011189

Seale, C. (1999). *The quality of qualitative research*. Sage Publications.

Shafiee Rad, H. (2024). Revolutionizing L2 speaking proficiency, willingness to communicate, and perceptions through artificial intelligence: A case of Speeko application. *Innovation in Language Learning and Teaching, 18*(4), 364–379. https://doi.org/10.1080/17501229.2024.2309539

Tapalova, O., & Zhiyenbayeva, N. (2022). Artificial intelligence in education: AIEd for personalised learning pathways. *Electronic Journal of E-Learning, 20*(5), 639–653. https://doi.org/10.34190/ejel.20.5.2597

Tavakoli, P., & Hunter, A.-M. (2018). Is fluency being 'neglected' in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research, 22*(3), 330–349. https://doi.org/10.1177/1362168817708462

Tavakoli, P., Nakatsuhara, F., & Hunter, A. M. (2020). Aspects of fluency across assessed levels of speaking proficiency. *The Modern Language Journal, 104*(1), 169–191. https://doi.org/10.1111/modl.12620

Thao, L. T., & Xuan, L. T. T. (2020). English language teaching reforms in Vietnam: EFL teachers' perceptions of their responses and the influential factors. *Innovation in Language Learning and Teaching, 16*(1), 29–40. https://doi.org/10.1080/17501229.2020.1846041

Tran, P. M., & Tanemura, K. (2020). English in Vietnam. *World Englishes, 39*(3), 528–541. https://doi.org/10.1111/weng.12489

Ushioda, E. (2011). Why autonomy? Insights from motivation theory and research. *Innovation in Language Learning and Teaching, 5*(2), 221–232. https://doi.org/10.1080/17501229.2011.577536

van den Berghe, R., Verhagen, J., Oudgenoeg-Paz, O., van der Ven, S., & Leseman, P. (2019). Social robots for language learning: A review. *Review of Educational Research, 89*(2), 259–295. https://doi.org/10.3102/0034654318821286

van Lier, L. (2005). Case study. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 195–208). Erlbaum. https://doi.org/10.4324/9780203836507

Van, L. T., & Duong, L. H. (2022). Exploring Vietnamese university students' perception of using mobile assisted language learning (MALL) to study English. *Ho Chi Minh City Open University Journal of Science – Social Sciences (Online), 12*(2), 152–163. https://doi.org/10.46223/HCMCOUJS.soci.en.12.2.2298.2022

Van Tuyen, L., & Loan, T. T. T. (2019). Factors affecting EFL students' willingness to communicate in speaking classes at the Vietnamese tertiary level. *International Journal of English, Literature and Social Science, 4*(2), 252–262. https://doi.org/10.22161/ijels.4.2.10

Wang, X., Liu, Q., Pang, H., Tan, S. C., Lei, J., Wallace, M. P., & Li, L. (2023). What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Computers & Education, 194*, 104703. https://doi.org/10.1016/j.compedu.2022.104703

Wang, X., Pang, H., Wallace, M. P., Wang, Q., & Chen, W. (2024). Learners' perceived AI presences in AI-supported language learning: A study of AI as a humanized agent from the community of inquiry. *Computer Assisted Language Learning, 37*(4), 814–840. https://doi.org/10.1080/09588221.2022.2056203

Xie, H., Chu, H.-C., Hwang, G.-J., & Wang, C.-C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. *Computers & Education, 140*, 103599. https://doi.org/10.1016/j.compedu.2019.103599

Yin, R. K. (2015). *Qualitative research from start to finish* (2nd ed.). Guilford Publications.

Yoshida, M. T. (2018). Choosing technology tools to meet pronunciation teaching and learning goals. *The CATESOL Journal, 30*(1), 195. https://doi.org/10.5070/B5.35971

Zechner, K., & Hsieh, C. (2024). Automated scoring and feedback for spoken language. In M. D. Shermis & J. A. Wilson (Eds.), *Routledge international handbook of automated essay evaluation* (pp. 141–160). Routledge. https://doi.org/10.4324/9781003397618-10

Zhao, D., Jablonkai, R. R., & Sandoval-Hernandez, A. (2024). Enhancing willingness to communicate in English among Chinese students in the UK: The impact of MALL with Duolingo and HelloTalk. *Journal of China Computer-Assisted Language Learning, 4*(1), 42–73. https://doi.org/10.1515/jccall-2023-0027

Zou, B., Liu, Q., Han, Y., Li, Z., & Zhang, W. (2023). Exploring students' acceptance of an artificial intelligence speech evaluation program for EFL speaking practice: An application of the Integrated Model of Technology Acceptance. *Computer Assisted Language Learning*, 1–26. https://doi.org/10.1080/09588221.2023.2278608

The Hidden Life of the Classroom

Jack C Richards[1]

University of Sydney

ABSTRACT

While digital technologies offer expanding opportunities for language learning beyond the classroom, the classroom itself remains a vital and complex site of language instruction. This article investigates the hidden dimensions of classroom life by examining the constructs of the *teacher self* and *learner self*. Drawing on research, classroom observations, and practitioner narratives, it explores how identity, beliefs, motivation, and agency shape engagement and interaction in language learning environments. Both the teacher self and the learner self are conceptualised in terms of identity, beliefs, and motivation. These constructs significantly influence classroom dynamics, pedagogical choices, and learning outcomes. The article advocates for reflective practices that enable educators to uncover and respond to these underlying dimensions, thereby enhancing teaching effectiveness and learner engagement. It offers practical insights for TESOL professionals seeking to better understand and support the evolving identities and motivations of both teachers and learners in diverse educational contexts.

Keywords: classroom dynamics, learner self, motivation, teacher identity, TESOL

[1] Correspondence: Jack C Richards: jack.richards@sydney.edu.au; https://orcid.org/0000-0001-9746-0852

ISSN 1834-4712 (Online).

DOI: 10.5281/zenodo.17588987

Citations: Richards, J. C. (2025). The hidden life of the classroom. *University of Sydney Journal of TESOL*, *4*, 85–112. https://doi.org/10.5281/zenodo.17588987

INTRODUCTION

Despite the growing availability of out-of-class learning opportunities through digital technologies, mobile devices, and media platforms, the classroom remains a central and enduring site for language instruction. Over the course of a career, a full-time teacher may spend between 30,000 and 40,000 hours in the classroom. Similarly, students typically accumulate approximately 900 instructional hours per year, amounting to over 11,000 hours between the ages of 5 and 18. These figures underscore the significance of classroom-based learning in shaping educational experiences. Language teaching and learning within the classroom are not solely cognitive processes; they are deeply embedded in social, emotional, and cultural dimensions. The classroom is a dynamic space where both teachers and learners bring their identities, beliefs, motivations, and histories into play. These personal and professional constructs influence how individuals engage in teaching and learning, as well as their perceptions of their roles in the classroom.

This article explores the *hidden life of the classroom* by examining how the constructs of the *teacher-self* and *learner-self* shape classroom engagement. It considers how identity, beliefs, and motivation influence pedagogical practices and learning outcomes, and argues for the importance of reflective inquiry in uncovering these often unseen dimensions of classroom life.

THE TEACHER SELF

The dimensions of teacher knowledge and practice explored here contribute to what can be called the *teacher self*—a complex and multidimensional construct encompassing a teacher's sense of identity, beliefs, values, emotions, experiences, motivation, professional knowledge, and personal history. These elements shape how teachers understand their role, how they perceive their students, the kind of teacher they aspire to be, and how they approach classroom instruction.

Sachs (2005) emphasises that teacher identity is not fixed but negotiated through experience and reflection. Bandura (2006) highlights the role of self-efficacy in shaping motivation and resilience in teaching. The distinction between the *open-self*, *secret-self*, *blind-self*, and *undiscovered-self* aligns with the Johari Window framework developed by Luft and Ingham (1955), which offers a lens for understanding self-awareness in interpersonal and professional contexts.

- The *open self* refers to aspects of a teacher's behaviour that are known both to the teacher and to others – that is, information that the teacher is willing and able to share with others. For example, 'I know that I often speak very quickly, and I sometimes ask students to signal to me when I am going too fast.'
- The *secret self* refers to information that is known to the teacher but not to others. For example, the teacher sometimes experiences stress and anxiety when teaching, but she does not want this to be obvious to others.
- The *blind self* refers to information that is known to others but not to the teacher. For example, the teacher may unintentionally spend more time with some students than others, resulting in unequal opportunities for all students to participate in the lesson.
- The *hidden self* refers to aspects of a teacher's identity and self-awareness that are unknown to both the teacher and to others. For example, neither the teacher nor others in the school may understand why the teacher is achieving lower-than-expected results with a new set of teaching materials.

In this article, I will focus on three aspects of the teacher self: teacher identity, teacher beliefs, and teacher motivation.

Teacher identity

We typically think of identity in relation to an individual's personal identity. *Personal identity* refers to unique and relatively stable features of a person's inner life that are performed or realised in contact and interaction with others. This image of the inner self, which a person presents to others, reflects features such as personality, age, gender, values, beliefs, life experience, self-image, or occupation, and depends on a person's role in an interaction—such as parent, son or daughter, friend, partner, customer, or student (Gee, 2000; Jenkins, 2008).

Teacher identity may reflect aspects of an individual's personal identity, but is primarily shaped by the nature of teaching itself. Teaching is a profession defined by the roles and functions teachers perform and the specialised knowledge and skills they bring to their work. As Pennington and Richards (2016) argue, teacher identity integrates personal, contextual, and professional factors and evolves through experience, reflection, and interaction within specific teaching contexts.

In the classroom, teachers enact a professional identity that may vary depending on the learners they teach. For instance, a teacher of young learners may adopt the role of storyteller or nurturer, while a teacher of adults may act as facilitator or motivator (Beijaard et al., 2004). These roles are not fixed but are shaped by pedagogical goals, institutional expectations, and the teacher's own beliefs and values. Teachers are expected to demonstrate attributes, such as commitment, subject knowledge, proficiency in English, and pedagogical skill. They are also expected to embody qualities such as warmth, confidence, enthusiasm, fairness, and emotional stability (Day et al., 2006). A teacher's professional identity is achieved and maintained through actions that may be both unconscious and strategic, as well as those that are consciously managed.

The teacher's proficiency in English is an important aspect of an English teacher's professional identity. A teacher whose English is not their dominant language may feel the need to maintain a strong command of English while teaching, which may require them to highlight different aspects of their professional identity than when teaching in their mother tongue.

> I think the main problem in my teaching is my English. I do not have a natural ability to handle the language, so I very often have to stop, interrupt myself, and find the words. It's embarrassing. (Danish teacher)

> I think my English proficiency is not good enough. I really want to do my best in every lesson. However, I often felt frustrated because I couldn't achieve my goals or meet the standards I had set before class. I always practised my English lessons before the class started because I did not want to lose face in front of my students. (Chinese student teacher)

Hence, more "identity work" may be needed to help the teacher realise the sense of academic authority that they have when teaching in their L1, and to be accepted by their students as a competent university lecturer or teacher.

TEACHER BELIEFS AND PRINCIPLES

Observing novice and experienced teachers was a regular aspect of my job when I worked as head of a large English department in a university. Hong Kong. Often, when I watched highly competent and experienced teachers teaching similar materials to similar classes of students, their classes functioned very differently. One reason for differences of this kind may be due to the different beliefs and

principles that individual teachers seek to realise in their teaching. Researchers refer to these as teachers' personal constructs of teaching, which shape how they approach their teaching. Shulman (1987) commented:

> The final source of the knowledge base [of teaching] … is the wisdom of practice itself, the maxims that guide or provide reflective rationalization for the practice of able teachers (p. 11).

What Shulman refers to as maxims are the working principles or beliefs that teachers consciously or unconsciously refer to as they teach. They are implicit theories about how students learn best and what makes for effective instruction. These can be inferred from observing how teachers conduct their lessons and from conversations with them about what they seek to achieve through their teaching. Wu and Richards (in press) examined Taiwanese EMI teachers' beliefs about the importance of English in their discipline and the type of language support they provided to their students. Significant differences were found between teachers' beliefs and practices in hard sciences, such as physics and engineering, and in soft subjects, such as history and education. Teachers in the hard sciences paid more attention to students' learning of content than to their use of English compared to teachers in the soft subjects. The former were also much more likely to use translanguaging in their teaching than soft subject teachers, who believed in the importance of providing an English-only language input to their students (See also Kao et al., 2021).

Teacher beliefs may be quite general (such as the idea that all lessons should connect to students' lives) or quite specific (such as the notion that peer error correction is more effective than teacher error correction). They may also be specific to a particular group of learners or to principles they seek to apply across all their teaching. An example of the former would be when a class contains students of very mixed levels, in which case the teacher seeks to realise these principles in her teaching:

- Vary group structures (pairs, small groups, whole class) to provide different learning contexts and opportunities for collaboration.
- Provide support and guidance to students who need it
- Set individualised learning goals for each student

A teacher I observed in Hong Kong described how her principles shaped her teaching:

> As the years go by, I feel less of a 'teacher' and increasingly, a 'guide' or 'conductor', trying to extract what is inside my students: trying to give them courage and develop trust between themselves so that their learning experiences will be faster, easier and fun in the process. With that philosophy of a more humanistic approach behind me, I always strive to create lessons that foster communication and cooperation among learners, where I take a back seat and they rely on each other more – a more student-centred lesson. So, naturally, a successful lesson would have incorporated all of this – me as a guide, an example, initially, someone they can rely on for help and where I can slowly dissolve into the background while they are discussing in English the task they are doing, and where I reappear only to help encourage, and apologise for the lesson being over.

Teachers' personal principles profoundly influence their teaching and are shaped by their experience as teachers, professional development, and their own experiences as learners. The following examples (simulated by ChatGPT 4.0) show how differences in two teachers' beliefs lead to very different instructional choices when they teach:

Teacher A: -- A Communicative Approach

Beliefs:

- Language is best learned through communication.
- Making mistakes is part of learning.
- Students learn more when they are active and engaged.

Classroom Practices:

- Uses pair and group work regularly (e.g., information gap activities, role plays).
- Focuses on fluency over grammatical accuracy during speaking tasks.
- Uses authentic materials like news clips, podcasts, and menus.
- Encourages student talk time—she speaks less, so students speak more.
- Gives formative feedback rather than correcting every error.

Teacher B: – A Form-Focused Approach

Beliefs:

- Accurate grammar and vocabulary are the foundation of language.
- Students need a clear structure and correction to improve.
- Learning is most efficient when guided step-by-step.

Classroom Practices:

- Begins lessons with explicit grammar instruction (e.g., rules of the past tense).
- Emphasises drills, repetition, and controlled practice before free use.
- Frequently uses textbooks and grammar exercises.
- Corrects spoken and written errors immediately.
- Uses tests and quizzes regularly to track progress.

The notion of teachers' beliefs and working principles needs to be kept in mind when you observe a teacher's class, because you may not be able to make sense of what happens during the lesson unless you clarify first with the cooperating teacher – the kinds of beliefs or principles that she seeks to realise in her teaching.

TEACHER MOTIVATION

Teacher motivation has been described as the energy that sustains and enhances teaching effectiveness, student engagement, and the overall educational environment (Han & Yin, 2016). It plays a crucial role in a teacher's job satisfaction, professional growth, and resilience in facing the challenges of the teaching profession (Klassen & Tze, 2014). Motivation influences not only instructional quality but also teacher retention, well-being, and the ability to adapt to changing educational demands. Some of its key attributes — commitment, self-esteem, agency, and self-efficacy — have been widely studied in recent literature.

Commitment refers to the teacher's personal engagement with teaching, the extent to which they have a sense of vocation, identify with and support the school's goals and practices and are willing to invest personal resources of time and energy to achieve excellence in teaching (Day et al., 2006).

> I am a teacher. But I am not simply a teacher. I am an English, social studies, and math teacher, a teacher of teachers, and a student of teachers. I believe in and am committed to a just society, equity of outcomes, ongoing dialogue with students, and professionalism, as well as maintaining professional competency. I also value inquiry-based communities, high expectations, and thoughtful practice (Hsieh, 2010, p. 1)

An Australian-Brazilian English teacher at an Australian university cites in conversations with colleagues how her commitment to self-improvement is an important part of her teacher identity:

> Actually, there is a bit of competition between the teachers regarding status; for example, the EAP teachers have a slightly higher status than the upper-intermediate teachers, and so on, even though that shouldn't be the case. As you know, I'm currently pursuing a PhD, and sometimes I can't help myself from saying 'in my research' in general staff room conversations, even though I know this isn't going to make anyone like me more. I guess it marks my status as a teacher who wants to improve themselves rather than one who is just getting by.
>
> I think I have a sense of joy from being a teacher. I really hope I can do more as their teacher. I want to help them improve their test results, and I also want to be their friend. I want to take care of them and support them.

Self-esteem refers to an individual's attitudes towards themselves and the extent to which they believe themselves to be successful, competent, and of value to others (Howard & Johnson, 2004). Positive self-esteem contributes to a teacher's social competence, enabling effective communication with students and colleagues, and provides emotional support and job satisfaction, giving the teacher feelings of confidence and strong coping skills. Self-esteem also relates to the value, status and importance a teacher attributes to language teaching as a profession.

> Whenever I have a conversation with someone, the peak of my pride and honour is when I introduce myself as a teacher. Unfortunately, many people take teachers for granted and don't value teaching as a profession. So, I try to be a positive representative of my profession.

Agency refers to the extent to which teachers can actively contribute to and manage change in their own teaching and professional development (Priestley et al., 2015). Rather than being the recipient of decisions and changes initiated by others, teacher agency is characterised by the teacher's ability to take ownership of their own learning and environment, and to set goals, develop curriculum, initiate change, and make decisions that affect the teacher's work and its conditions.

> I like to try out new things in my teaching. To do so, I do take charge of my own learning and try to learn a new language using my own tips and techniques. This is a great way to experience what my learners go through on their language learning journey. Besides, by using this strategy, I get a lot of ideas of what works best for me as a basis for my future plans.

Self-efficacy refers to the teacher's view of his or her own effectiveness – that is, the ability to perform well as a teacher of English, to achieve their goals and potential, to maintain their commitment to teaching despite difficulties they may encounter and to provide support for students' learning (Tschannen-Moran & Hoy, 2001). It is linked to positive teaching experiences. Teacher self-efficacy is linked to positive teaching experiences, such as observing students' progress and receiving positive feedback from students and others, which contributes to the teacher's sense of identity as a competent and successful educator. This, in turn, contributes to their sense of agency as well as their commitment to teaching.

> I believe that to be effective, I have to inspire my students with the idea that their success in learning English will prepare them well for the future. This will also require more independent learning to achieve their goal. So, putting

the students on the right path is the key to my effectiveness and success as a teacher.

THE LEARNER SELF

By analogy with the notion of the *teacher self*, the *learner self* refers to how students perceive themselves as learners, encompassing their beliefs, identities, goals, motivations, emotions, and abilities. This concept encompasses learners' perceptions of language learning, their responses to the learning experience, their understanding of their role in classroom learning, and their management of both cognitive and emotional aspects of the learning process. Learners' perceptions of their own abilities and roles in the classroom shape their engagement, performance, confidence, and participation, which, in turn, influence their commitment to learning and their learning outcomes (Mercer, 2011; Dörnyei, 2009).

The learner self also involves the disposition learners have towards classroom learning and the kind of participation they favour. Research has shown that learners' self-concept—how they perceive their competence and identity as language users—is a central psychological construct that affects motivation and engagement (Mercer, 2011). For example, Mercer argues that self-concept is domain-specific and relatively stable, influencing learners' willingness to engage and persist in language learning tasks. Similarly, Sahinkarakas and Inozu (2017) found that learners' future L2 self-guides and vision significantly predict their ideal L2 self, which in turn drives motivation and classroom participation.

Moreover, learner identity is dynamic and socially constructed, shaped by intercultural experiences and classroom interactions. Poststructuralist perspectives in second language acquisition (SLA) emphasise that learners are active agents whose identities evolve through engagement with diverse linguistic and cultural contexts (Norton, 2013). This evolving identity influences learners' emotional responses—such as anxiety or enjoyment—and their willingness to communicate, which are critical to successful language learning (Dörnyei & Ryan, 2015).

Ultimately, emotional engagement and self-regulation are crucial components of the learner's self. Studies have shown that learners with high levels of grit and self-regulation exhibit stronger classroom engagement and better academic outcomes (Duckworth et al., 2007; Ushioda, 2011). Motivation and engagement, as

behavioural manifestations of the learner self, are influenced by learners' visions of themselves and their perceived autonomy, competence, and relatedness within the learning environment (Ryan & Deci, 2000). For example:

- *Relaxed or anxious*: Some learners may find the language class an opportunity to relax and enjoy themselves. For others, it may arouse anxiety, and they may be reluctant to participate.

- *Risk-taking or risk-avoiding*: Some students are not afraid to try new things, even if they may not be very successful. Others avoid activities that involve risks.

- *Playful or serious*: Some students enjoy activities that have a playful and fun element. Others think learning is serious business and may dislike such activities.

- *Fuzzy focused or black and white*: Some learners don't need clear outcomes for activities. Others prefer an unambiguous 'black-and-white' outcome.

- *Confident or insecure*: Some students have confidence in their abilities, while some may be less sure of their abilities.

As with the discussion of teacher self, I will consider the notion of learner self in relation to learner identity, learner beliefs, and learner motivation.

Learner identity

Learner identity involves the beliefs, values, and commitments an individual holds toward being a language learner (as distinct from other types of learners), as well as how others view their status as a language learner. It reflects the nature of interactions and their social context, and is considered an interactional achievement shaped by intent, power, and self-image (Norton, 2013; Block, 2007). This identity can significantly influence learners' attitudes and motivation toward English and language learning, their participation in classroom instruction, and their use of English beyond the classroom (Dörnyei & Ryan, 2015).

Learner identity is not static; it is dynamic, socially constructed, and often negotiated through interaction with others in various communities of practice (Pavlenko & Blackledge, 2004; Wenger, 1998). Learners bring particular identities to the classroom—such as being a mother, engineer, teenager, or

businessperson—but within the language classroom, these identities are often redefined, for example, as beginning, intermediate, or advanced-level learners. This redefinition is shaped by both internal perceptions and external positioning, including how learners are perceived by peers and teachers (Duff, 2012).

Poststructuralist perspectives in second language acquisition (SLA) emphasise that identity is multiple, fluid, and a site of struggle, changing over time and across contexts (Norton, 2013). Learners continuously negotiate their sense of self in relation to the social world, and this negotiation affects their investment in language learning. The concept of *investment*—as distinct from motivation—captures the socially and historically constructed relationship learners have with the target language and their desire to gain symbolic and material resources through language learning (Norton, 1995).

Furthermore, learner identity is closely tied to classroom dynamics and broader sociocultural contexts. Identity development is influenced by learners' sense of agency, their perceptions of affordances in the learning environment, and mismatches between imagined and practised communities (Kanno & Norton, 2003). In intercultural and study abroad contexts, identity conflicts—such as being positioned as a "foreigner"—can impact learners' engagement and acquisition of linguistic competence (Block, 2007). These experiences highlight the importance of viewing learners as active agents whose identities are co-constructed through interaction and shaped by power relations and cultural narratives (Pavlenko & Blackledge, 2004).

A learner's identity is defined by their interactions with the teacher and others, both in and out of the classroom. Lam (2000) documented an interesting identity transformation of a Chinese immigrant teenager through his electronic textual experiences. Whereas Almon was largely stigmatised as a low-achieving English as a second language student and struggled to develop English literacy in his school environment, his engagement with written English on the Internet with a transnational peer group in English allowed him to construct a more confident identity in this web-based context and to develop a sense of belonging to a global English-speaking community.

For some, the classroom may represent a site of struggle. Remaining silent is one way of protecting the learner's identity. A learner may be a highly motivated language learner but may nevertheless have little investment in the language practices of a given classroom, which may, for example, be racist, sexist, elitist, or

homophobic. Thus, despite being highly motivated, a learner could be excluded from classroom language practices and, over time, positioned as a 'poor' or unmotivated language learner.

Learner beliefs

Learner beliefs are the preconceived ideas or notions learners hold about different aspects of language and language learning. Learners may come to class with pre-formed ideas about how best to learn vocabulary or grammar, the kinds of people who do best at language learning, the importance of translation, the significance of a foreign accent, and the value of practice, repetition, error correction, and group work (Barcelos & Kalaja, 2011; Bernat & Gvozdenko, 2005). These beliefs are often shaped by prior learning experiences, cultural and educational backgrounds, and observations of teachers and peers. For example, some learners may hold stereotypical views such as 'girls are better at languages than boys', which reflect broader societal beliefs and can influence classroom dynamics and learner expectations (see also Bernat & Gvozdenko, 2005).

Learners may also have fixed ideas about the language they are studying. In the case of English, some may believe that certain varieties—such as British or American English—are superior to others, or that English has more or less grammar than other languages they have studied (Zhong, 2022). These beliefs can affect learners' attitudes toward different accents and dialects, and their openness to diverse forms of English. Additionally, learners set different goals for themselves in learning English. For some, native-like mastery is the ultimate goal, while others prioritise communicative competence and view localised use of English—including pronunciation features of their mother tongue—as a valid and culturally significant form of expression (Nguyen et al., 2014).

Importantly, learner beliefs are not static. They can evolve over time through reflection, interaction, and exposure to new learning contexts. Understanding and addressing learners' beliefs is crucial for teachers, as these beliefs influence learners' motivation, strategy use, and engagement with language tasks. Teachers and learners often hold differing beliefs about language learning, which can lead to curriculum misalignment. For instance, a teacher may value group work, while students may doubt their ability to learn through interactions with peers who have limited English proficiency. Or when students come from a culture where rote learning and memorisation are widely used and considered useful strategies for learning English, whereas the teacher does not believe in their value and tries to

discourage learners from using them. Students may also have very specific views and expectations about what they expect of a teacher, as we see in these comments from Australian students:

> A good teacher is someone who can persuade other people to express themselves and bring out the creativity in kids. I think everyone has creativity they don't use, and it's the teacher's ability to bring that out in the person and persuade them to use it.
>
> The ultimate thing for a teacher is the ability and willingness to sit down and talk to students, because you find that some English teachers (they're in a minority) aren't willing to talk to their students on a level…; you get people who talk down which is the worst thing for the writing of English because you need to be able to discuss things absolutely straight.

Leaners may also have different views about the goals and purpose of different kinds of classroom activities. In the case of reading, for example, in the same class of young learners, students may have different beliefs about the nature and purpose of reading:

- *Reading is saying words correctly*: The focus is on saying words aloud. Reading is a performance of calling out words.
- *Reading is schoolwork*: Reading is just another assignment to be completed before moving on to something more enjoyable. It is not something they would normally choose to do on their own.
- *Reading is a source of status*: It is an activity to be announced and performed in front of the class.
- *Reading is a way to learn things*: Reading is studying, and they choose to read things that contain interesting and important information.
- *Reading is a private pleasure*: Children read things that have personal meaning for them.
- *Reading is a social activity*: Reading is a shared activity conducted collaboratively in pairs or groups. It is a source of pleasure with friends.

When classes consist of students from different cultures or when the teacher does not share the same linguistic and cultural background as the students, some students may find that the norms for classroom learning in the English lesson are different from those in their own culture, reflecting different beliefs about the nature of good teaching, as we see in these comments.

> The trouble with Chinese teachers is that they've never done any real teacher-training courses, so they don't know how to teach. All they do is follow the book. They never give us any opportunity to talk. How in the world do they expect us to learn? (Australian student studying in China)
>
> Australian teachers are very friendly but they can't teach very well. I never know where they're going – there's no system and I just get lost. Also, they're often poorly trained and lack a thorough grasp of their subject. (Chinese student studying in Australia, Brick, 2004)

Many Asian learners are also familiar with a teaching approach influenced by Confucian educational values and may initially be resistant to the learner-centred approach used in other countries. They may tend to view teachers as authority figures, which differs from how they are perceived in Western culture. When I started teaching in the US after many years in Asia, I was surprised that many students used my first name without being invited to do so. Whereas some former students of mine in Thailand who are now university professors still insist on using the honorific ajarn and feel uncomfortable calling me by my first name.

Learners may have to accommodate a different set of assumptions about the nature of classroom communication from those operating in their home culture. This may mean adopting new ways of interacting that reflect different norms for interpersonal communication, particularly in the pragmatic domain—such as differences in degrees of taciturnity, including the keeping of one's thoughts and emotions to oneself, as well as issues of control, reserve, reticence, self-restraint, and communicativeness (Scollon et al., 2012; Yates & Nguyen, 2012). These differences are often rooted in culturally specific expectations about appropriate behaviour and communication, which can affect learners' comfort and participation in classroom discourse.

The learner's view of what constitutes a "good learner" may also differ from the teacher's, leading to resistance toward classroom behaviours that conflict with their cultural identity. For example, learners from cultures that value humility—such as Japanese or Chinese students—may avoid speaking English in front of their

classmates to prevent being perceived as "showing off" (Cortazzi & Jin, 1996; Jin & Cortazzi, 2006). This reluctance to speak publicly is often discussed in terms of "willingness to communicate" (WTC), a concept widely studied in Chinese EFL contexts. Research shows that Chinese learners may experience speaking anxiety and reticence due to cultural norms that discourage overt self-expression, especially in group settings (Peng, 2012; Wen & Clement, 2003).

Such resistance is not merely behavioural but reflects deeper identity negotiations. Learners may resist classroom practices they perceive as misaligned with their cultural values or personal identity, particularly when power dynamics and classroom positioning challenge their sense of self (Norton, 2013). Creating inclusive classroom environments that affirm learners' cultural identities and respect diverse communication styles can help mitigate these tensions and foster more meaningful engagement (Kramsch, 1993).

Motivation

Motivation is often described as the driving force of second language learning, and without a strong desire to learn English, learners are unlikely to invest the time and effort required (Dörnyei & Ushioda, 2011; Gardner, 2010). Students with a strong interest in learning English may view it as having "capital" or potential value. Their motivation may reflect different values they associate with proficiency in English, such as academic achievement, future job opportunities, communication through social media, cross-cultural friendships, participation in popular culture through movies, television, and pop music, international travel, the opportunity to develop an international outlook, and the prestige associated with bilingual fluency in English (Norton, 2013; Ushioda, 2017).

Dörnyei (2009) proposed that learners can be motivated by an imagined view of themselves—a vision of the person they would like to become—and that mastery of English is a core component of this possible self, functioning as a powerful motivational tool. His L2 Motivational Self System includes three components: the ideal L2 self (who the learner wants to become), the ought-to L2 self (who others expect the learner to become), and the L2 learning experience (the learner's immediate learning environment). These components have been shown to significantly predict learners' intended effort and engagement in language learning.

Classroom learning conditions can do a great deal to maintain learners' motivation. Lessons that are motivating for students contain purposeful activities at an

appropriate level; they are not overloaded with content; they promote learner involvement and engagement; and they give learners a sense of success rather than failure (Dörnyei & Csizér, 1998). This reflects what is sometimes referred to as the **Goldilocks Rule**, which suggests that peak motivation occurs when learners are working on tasks that are "just right"—not too easy and not too hard, but at the edge of their current abilities. In a similar sense, both Kagan (1990) and Treffers-Daller (2024) use the term *Goldilocks* to describe the need for the appropriate amount of something—whether in the complexity of ideas or the proportion of language use in learning—highlighting the importance of balance for effective engagement and learning.

Self-efficacy

Self-efficacy refers to the learner's view of their own effectiveness—that is, their belief in their ability to succeed in learning English, achieve their goals and potential, and maintain their commitment to learning English despite difficulties they may encounter (Bandura, 1997). Self-efficacy is closely linked to the learner's positive experiences in learning English and plays a critical role in shaping their motivation, persistence, and performance (Chang et al., 2024). Learners with high self-efficacy are more likely to engage actively in language learning tasks because they believe their efforts will lead to success. They are more willing to take risks, such as speaking in class or using the language outside the classroom, and they tend to persist longer when faced with challenges (Cai, 2024).

Moreover, self-efficacy is not a fixed trait—it can be developed through mastery experiences, observing others succeed (vicarious experiences), receiving encouragement (verbal persuasion), and managing emotional states (Bandura, 1986). In second language contexts, self-efficacy has been shown to be a strong predictor of performance across language skills, including speaking, writing, and listening. It also mediates the relationship between learners' motivation and their actual achievement, making it a key factor in successful language acquisition.

> Whenever I saw a tourist, I would try my best not to be shy and talk to them. This talking to strangers happened mostly on Yahoo Messenger when chatting with people around the world. (Hamed)

> But I think I learned more through communicating with foreigners. I often approached foreigners I met to speak with them. (Vida)

Chilian student Danilo has started to learn German, and described how he makes use of Instagram to learn through interaction with other learners of German:

> I created an Instagram account in which I only followed profiles that posted content for learning German and many other topics, all in German. By scrolling through social media in German every day, I am absorbing German vocabulary and grammar. At first, it was quite tough, but after about a month and a half, I started being able to understand most of what I read, and I can also understand the structure of the language as well. I also complement this with web research and the use of ChatGPT to learn basic conversation and phrases.

High self-efficacy helps learners persist through difficulties, such as understanding grammar or pronunciation, and contributes to resilience in the face of setbacks, reducing the likelihood of giving up (Bandura, 1997). Learners with strong self-efficacy are more likely to use effective learning strategies, such as seeking feedback, practising outside of class, and employing metacognitive strategies, such as planning, monitoring, and evaluating their learning (Chang et al., 2024). These learners tend to approach challenges with confidence and are more proactive in managing their learning.

When learners believe they can succeed, they experience less language anxiety, especially in speaking and listening contexts. Self-efficacy has been shown to buffer against anxiety by fostering a sense of control and competence in communication tasks (Wang, 2024). Research consistently shows that self-efficacy is a strong predictor of actual performance in second language learning—often more so than aptitude or prior achievement (Mills, Pajares, & Herron, 2007; Pérez Navío et al., 2024). This makes self-efficacy a critical factor in both the cognitive and emotional dimensions of language acquisition.

Agency

Agency refers to the extent to which learners can actively contribute to and manage change in their own learning, act independently, and make choices about their learning (Duff, 2012; Mercer, 2011). It positions learners not as passive recipients of knowledge but as active participants in the learning process. This is reflected in their motivation and engagement: learners with a strong sense of agency are more likely to set goals, persist through challenges, and take initiative, resulting in more sustained engagement with the language (Ushioda, 2011).

Learners with agency often create personalised learning paths by seeking methods and resources that suit their interests and learning styles—for example, listening to music, watching TV shows, or engaging in conversation to develop language skills (Lamb, 2017). They look for opportunities beyond the classroom and develop a sense of autonomy, often acquiring skills in self-monitoring, reflection, and self-correction (Benson, 2013). This autonomy enables them to take responsibility for their own progress rather than relying solely on teachers, fostering a deeper, more meaningful learning experience (Little, 2007).

Danilo, the Chilean student I referred to earlier, experienced his introduction to English in a very similar way to the examples of traditional classroom teaching cited above. Finding that he was learning very little from it, he took matters into his own hands. For him, the three most useful activities for learning English were watching TV shows and movies, watching YouTube clips, and learning songs. On the value of watching movies, he commented:

> TV and movies help the most because there's plenty of dialogue between characters, and they usually speak correctly. If they have a different way of speaking, it is always noticeable which group the person comes from, which helps learning as well.

Danilo added:

> In Chile, my country, English taught in school is usually not useful at all, unless you go to an English private school, because you can be surrounded by it every day, all day, but other than that, even in an English after-school Institute, it is very likely that you come out of it knowing very little. Everyone I know here in Chile who speaks English, including myself, has learned on their own by immersing themselves in the language and by being willing to learn. I think what's the most important and helped me the most was that I would speak with myself, or try to think in English and use the words I learned for expressing things that happened around me, even though if I didn't have the need to, or I wasn't talking to anybody, so that the language would become a part of me.

CONCLUSIONS

We could say that the person you see at the tennis club becomes a different person when they step into the classroom as a teacher—their *teacher self* takes over. Similarly, the student you see enjoying a movie with friends becomes a different person when they enter the classroom—their *learner self* emerges. These shifts reflect the dynamic, context-dependent nature of identity, in which individuals enact different selves depending on the social setting and expectations (Gee, 2025; Norton, 2013).

Observing the lessons of experienced teachers is a regular activity in teacher education programs, and many forms of lesson observation have been developed to guide the process. These often include categories of teacher and learner performance such as the procedures the teacher used to scaffold lessons, the kinds of questions teachers asked, the amount of teacher talk and student talk, how group work was set up, the kinds of feedback provided, and the learners' use of English during the lesson. To ensure objectivity, observation forms and protocols generally focus on low-inference categories—observable and quantifiable actions that do not require interpretation or subjective judgment (Barkhuizen, 2016).

However, many of the features discussed above—such as beliefs, motivations, and identity—are high-inference aspects of the teacher and learner self. These are not easily captured through a single lesson observation. As Gee (2015) and Olsen et al. (2023) argue, identity is enacted through language, behaviour, and values, and

often includes elements that may be open, secret, blind, or undiscovered. Understanding these deeper dimensions requires sustained engagement with classroom life and reflective practice over time.

For teachers, exploring and understanding the nature and impact of the teacher self can occur through critical reflection, self-observation (e.g., video analysis), peer observation, case studies, reflection on critical incidents, and group-based reflective activities (Farrell, 2015; Zarrabi & Mohammadi, 2024). These practices help teachers uncover the values, beliefs, and motivations that shape their professional identity and teaching practice. Similarly, learners can reflect on their identities, beliefs, and learning experiences through narrative writing, classroom discussions, and the construction of personal stories about their journey as language learners (Miyahara, 2015). Such reflective practices foster greater self-awareness and agency, enabling both teachers and learners to engage more meaningfully in the educational process.

ACKNOWLEDGEMENT

I am grateful to Aek Phakiti and Keith Cheng Lin for comments and suggestions in the preparation of this paper.

AUTHOR

Jack C. Richards has had an active career in the Asia-Pacific region, having held chair professorships at the University of Hawaii, the City University of Hong Kong, and the University of Auckland. He is currently an honorary professor in the Sydney School of Education and Social Work, the Faculty of Arts and Social Sciences at the University of Sydney. He has also had a long association with the Regional Language Centre in Singapore. He has written over 150 books and articles on language teaching as well as many widely used classroom texts. He has been awarded an honorary Doctor of Literature by Victoria University of Wellington for his services to education and the arts, and has received the Arts Foundation of New Zealand Award for Patronage. The International TESOL organisation honoured Jack Richards as one of the 50 TESOL specialists worldwide to have made a significant impact on language teaching in the last 50 years. Website: www.professorjackrichards.com

REFERENCES

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman.

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science*, *1*(2), 164–180. https://doi.org/10.1111/j.1745-6916.2006.00011.x

Barcelos, A. M. F., & Kalaja, P. (2011). Introduction to beliefs about SLA revisited. *System, 39*(3), 281–289. https://doi.org/10.1016/j.system.2011.07.001

Barkhuizen, G. (2016). Language teacher identity research: An introduction. In G. Barkhuizen (Ed.), *Reflections on language teacher identity research* (pp. 1–11). Routledge. https://doi.org/10.4324/9781315643465-5

Benson, P. (2013). *Teaching and researching autonomy in language learning* (2nd ed.). Routledge. https://doi.org/10.4324/9781315833767

Bernat, E., & Gvozdenko, I. (2005). Beliefs about language learning: Current knowledge, pedagogical implications, and new research directions. *TESL-EJ*, *9*(1), 1–21. Retrieved from https://eric.ed.gov/?id=EJ1065832

Block, D. (2007). *Second language identities*. Continuum.

Brick, J. (2004). *China: A handbook in intercultural communication*. NCELTR, Macquarie University.

Chang, P., Zhou, L., & Zhang, L. J. (2024). The roles of task-specific self-efficacy and enjoyment in EFL learners' development of speaking CALF: A longitudinal study. *System*, 127, 103539. https://doi.org/10.1016/j.system.2024.103539

Cortazzi, M., & Jin, L. (1996). Cultures of learning: Language classrooms in China. In H. Coleman (Ed.), *Society and the language classroom* (pp. 169–206). Cambridge University Press.

Day, C., Kington, A., Stobart, G., & Sammons, P. (2006). *The personal and professional selves of teachers: Stable and unstable identities*. *British*

Educational Research Journal, *32*(4), 601–616. https://doi.org/10.1080/01411920600775316

Dörnyei, Z. (2009). *The L2 motivational self system*. In Z. Dörnyei & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9–42). Multilingual Matters.

Dörnyei, Z., & Csizér, K. (1998). Ten commandments for motivating language learners: Results of an empirical study. *Language Teaching Research*, *2*(3), 203–229. https://doi.org/10.1177/136216889800200303

Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. Routledge. https://doi.org/10.4324/9781315779553

Dörnyei, Z., & Ushioda, E. (2011). *Teaching and researching motivation* (2nd ed.). Routledge. https://doi.org/10.4324/9781315833750

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087

Duff, P. A. (2012). Identity, agency, and second language acquisition. In A. Mackey & S. M. Gass (Eds.), *The Routledge handbook of second language acquisition* (pp. 410–426). Routledge.

Farrell, T. S. C. (2015). *Reflective language teaching: From research to practice*. Bloomsbury.

Gardner, R. C. (2010). *Motivation and second language acquisition: The socio-educational model*. Peter Lang.

Gee, J. P. (2000). Identity as an analytic lens for research in education. *Review of Research in Education*, *25*(1), 99–125. https://doi.org/10.3102/0091732X025001099

Gee, J. P. (2015). *Social linguistics and literacies: Ideology in discourses* (5th ed.). Routledge. https://doi.org/10.4324/9781315722511

Gee, J.P. (2025). How to Do Discourse Analysis: A Toolkit (3rd ed.). Routledge. https://doi.org/10.4324/9781003537991

Han, J., & Yin, H. (2016). *Teacher motivation: Definition, research development and implications for teachers*. *Cogent Education*, *3*(1), 1217819. https://doi.org/10.1080/2331186X.2016.1217819

Howard, S., & Johnson, B. (2004). *Resilient teachers: Resisting stress and burnout*. *Social Psychology of Education*, *7*(4), 399–420. https://doi.org/10.1007/s11218-004-0975-0

Hsieh, B. Y.-C. (2010). *Exploring the complexity of teacher professional identity* (Doctoral dissertation, University of California, Berkeley). eScholarship. https://escholarship.org/uc/item/9406p4sb

Jenkins, R. (2014). *Social identity* (4th ed.). Routledge. https://doi.org/10.4324/9781315887104

Jin, L., & Cortazzi, M. (2006). Changing practices in Chinese cultures of learning. *Language, Culture and Curriculum*, *19*(1), 5–20. https://doi.org/10.1080/07908310608668751

Kagan, D. M. (1990). Ways of evaluating teacher cognition: Inferences concerning the Goldilocks principle. *Review of Educational Research, 60*(3), 419–469. https://doi.org/10.3102/00346543060003419

Kanno, Y., & Norton, B. (2003). Imagined communities and educational possibilities: Introduction. *Journal of Language, Identity & Education*, *2*(4), 241–249. https://doi.org/10.1207/S15327701JLIE0204_1

Kao, S.-M., Tsou, W., & Chen, F. (2021). Translanguaging Strategies for EMI Instruction in Taiwanese Higher Education. In W. Tsou & W. Baker (Eds.), *English-Medium Instruction Translanguaging Practices in Asia* (pp. 81–101). Springer Nature Singapore. https://doi.org/10.1007/978-981-16-3001-9_5

Klassen, R. M., & Tze, V. M. C. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review*, *12*, 59–76. https://doi.org/10.1016/j.edurev.2014.06.001

Kramsch, C. (1993). *Context and culture in language teaching*. Oxford University Press.

Lam, W. S. E. (2000). L2 literacy and the design of the self: A case study of a teenager writing on the Internet. *TESOL Quarterly*, *34*(3), 457–482. https://doi.org/10.2307/3587739

Lamb, M. (2017). The motivational dimension of language teaching. *Language Teaching, 50*(3), 301–346. https://doi.org/10.1017/S0261444817000088

Little, D. (2007). Language learner autonomy: Some fundamental considerations revisited. *Innovation in Language Learning and Teaching*, *1*(1), 14–29. https://doi.org/10.2167/illt040.0

Luft, J., & Ingham, H. (1955). *The Johari window: A graphic model for interpersonal relations*. University of California Western Training Lab.

Mercer, S. (2011). *Towards an understanding of language learner self-concept*. Springer.

Mills, N., Pajares, F., & Herron, C. (2007). Self-efficacy of college intermediate French students: Relation to achievement and motivation. *Language Learning*, *57*(3), 417–442. https://doi.org/10.1111/j.1467-9922.2007.00421.x

Miyahara, M. (2015). *Emerging self-identities and emotion in foreign language learning: A narrative-oriented approach*. Multilingual Matters.

Nguyen, H. T., Warren, W., & Fehring, H. (2014). Factors affecting English language teaching and learning in higher education. English Language Teaching, *7*(8), 94–105. https://doi.org/10.5539/elt.v7n8p94

Norton, B. (1995). Social identity, investment, and language learning. *TESOL Quarterly*, *29*(1), 9–31. https://doi.org/10.2307/3587803

Norton, B. (2013). *Identity and language learning: Extending the conversation*. Multilingual Matters.

Olsen, B., Buchanan, R., Hewko, C. (2023). Recent Trends in Teacher Identity Research and Pedagogy. In I. Menter (ed.), *Palgrave Handbook of Teacher*

Education Research. Palgrave Macmillan. https://doi.org/10.1007/978-3-031-16193-3_80

Pavlenko, A., & Blackledge, A. (Eds.). (2004). *Negotiation of identities in multilingual contexts*. Multilingual Matters.

Peng, J. E. (2012). Towards an ecological understanding of willingness to communicate in EFL classrooms in China. *System*, *40*(2), 203–213. https://doi.org/10.1016/j.system.2012.02.002

Pennington, M. C., & Richards, J. C. (2016). *Teacher identity in language teaching: Integrating personal, contextual, and professional factors*. RELC Journal, *47*(1), 5–23. https://doi.org/10.1177/0033688216631219

Priestley, M., Biesta, G., & Robinson, S. (2015). *Teacher agency: An ecological approach*. Bloomsbury Publishing.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sachs, J. (2005). Teacher education and the development of professional identity: Learning to be a teacher. In M. Kompf & P. Denicolo (Eds.), *Connecting policy and practice: Challenges for teaching and learning in schools and universities* (pp. 5–21). Routledge. https://doi.org/10.4324/9780203012529

Sahinkarakas, S., & Inozu, J. (Eds.). (2017). *The role of the self in language learning*. Cambridge Scholars Publishing.

Scollon, R., Scollon, S. W., & Jones, R. (2012). *Intercultural communication: A discourse approach* (3rd ed.). Blackwell.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–22. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Treffers-Daller, J. (2024). Unravelling translanguaging: A critical appraisal. *ELT Journal, 78*(1), 64–73. https://doi.org/10.1093/elt/ccad058

Tschannen-Moran, M., & Hoy, A. W. (2001). *Teacher efficacy: Capturing an elusive construct*. Teaching and Teacher Education, *17*(7), 783–805. https://doi.org/10.1016/S0742-051X(01)00036-1

Ushioda, E. (2011). *Language learning motivation, self and identity: Current theoretical perspectives*. *Computer Assisted Language Learning*, *24*(3), 199–210. https://doi.org/10.1080/09588221.2010.538701

Ushioda, E. (2017). The impact of global English on motivation to learn English in the European context. *Language Teaching*, *50*(3), 425–437. https://doi.org/10.1111/modl.12413

Wang, J. (2024). The relationship between foreign language anxiety and self-efficacy: A study in Chinese college students' English learning. *Lecture Notes in Education Psychology and Public Media*, *63*(1), 157–164. https://doi.org/10.54254/2753-7048/63/20240961

Wen, W. P., & Clement, R. (2003). A Chinese conceptualisation of willingness to communicate in ESL. *Language, Culture and Curriculum*, *16*(1), 18–38. https://doi.org/10.1080/07908310308666654

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press. https://doi.org/10.1017/CBO9780511803932

Wu, C-h, & Richards, J. C. (in press). Taiwanese EMI teachers' views of the role of language support for students in their disciplines. *Iranian Journal of Language Teaching Research*

Yates, L., & Nguyen, N. T. (2012). Beyond a discourse of deficit: The meaning of silence in the classroom. *The International Journal of Learning*, *18*(6), 1–12. https://eric.ed.gov/?id=EJ1003823

Zarrabi, M., & Mohammadi, M. (2024). Reflective teaching as an agent of reform in teacher education: Inhibitors, motivators, and tools. *International Journal of Educational Reform*, *33*(4), 503–521. https://doi.org/10.1177/10567879241255312

Zhong, Q. M. (2022). Learners' beliefs about English varieties: Implications for teaching English as an international language. In S. Li, P. Hiver, & M. Pap,

Routledge handbook of second language acquisition and individual differences (pp. 13–27). Routledge. https://doi.org/10.4324/9781003270546-18

Reimagining EAP in the Age of GenAI: Innovation, Integrity, and the Future of Assessment

Phuong Tran[1]

Monash College

Cara Dinneen

Macquarie University College

ABSTRACT

English for Academic Purposes (EAP) programs play a critical role in preparing students with English as an additional language for academic success in Australian universities. The internationalisation of higher education in Australia, driven by increased global mobility and shifting student demographics, has transformed the landscape of teaching, learning, and assessment in higher education. As universities adapt to the demands of a diverse and digitally connected student body, they must respond not only to linguistic and cultural diversity but also to the pedagogical implications of multimodal learning, digital collaboration, and the rise of generative artificial intelligence (GenAI). This article presents a dual-track assessment regime at two Australian university English language centres, Macquarie University College and Monash University English Language Centre,

[1] Correspondence: Phuong Tran: phuong.tran@monashcollege.edu.au; https://orcid.org/0009-0006-3429-5195; Cara Dinneen: cara.dinneen@mq.edu.au; https://orcid.org/0000-0002-2497-565X

ISSN 1834-4712 (Online).

DOI: 10.5281/zenodo.17590578

Citation: Tran, P., & Dinneen, C. (2025). Reimagining EAP in the age of GenAI: Innovation, integrity, and the future of assessment. *University of Sydney Journal of TESOL*, *4*, 113–135. https://doi.org/10.5281/zenodo.17590578

which offer high-stakes EAP programs for direct entry to university. The centres combine secure, invigilated checkpoints for progression decisions with guided, AI-assisted assessment for learning, aligned with regulatory expectations for direct entry programs. Drawing on policy analysis from the centres, the article contributes a practice-informed illustration of how EAP programs can integrate AI literacy, feedback literacy, and genre pedagogy while preserving credible evidence of unassisted performance. Emerging assessment designs and the application of GenAI in assessment development and feedback are discussed, alongside key risks and mitigation strategies. The paper concludes with implications for EAP programs that aim to equip students to navigate the complexities of contemporary higher education and make meaningful contributions to global academic communities.

Keywords: academic integrity, assessment, English for Academic Purposes, English language assessment, generative AI, target language use domain, higher education

INTRODUCTION

The increased global mobility of recent decades has led to the internationalisation of higher education (HE) in Australia, with students from diverse cultural backgrounds now forming part of student communities nationwide. This shifting demographic, alongside developments in communication technologies, generative artificial intelligence (GenAI) and an increasingly competitive global marketplace, has played a pivotal role in reshaping learning and teaching practices. Even prior to COVID-19, universities were witnessing the expansion of multimodality, interactivity, and collaboration in pedagogical approaches, as educators sought ways to integrate, engage, and socialise diverse cohorts of domestic and international students (Arkoudis et al., 2013).

In preparation for tertiary studies in Australia, international students with English as an additional language (EAL) can demonstrate the required level of English language proficiency by successfully completing an English for academic purposes (EAP) program at a registered English Language Intensive Courses for Overseas Students (ELICOS) centre. EAP programs are designed to prepare students from diverse cultural and educational backgrounds for their studies while ensuring they develop the requisite level of English proficiency. Programs run between 5 and 25 weeks, and on successful completion, EAL students are deemed to have sufficient language skills to undertake further studies in higher education or vocational education and training.

To ensure relevance and authenticity in their program design, EAP program developers monitor learning and assessment trends in the target language use (TLU) domain, specifically in universities. As pedagogical practices continue to evolve in response to shifting student demographics, technological advancements, and global academic standards, developers seek to align EAP curricula with the expectations of HE institutions. The recent proliferation of GenAI tools has further complicated this landscape, raising new questions about academic integrity, authorship, and the role of AI-assisted writing in student work. In response, EAP programs are increasingly tasked with not only refining language instruction but also embedding academic literacies, digital and AI literacies, ethical considerations, and assessment methods that reflect current university practices. By doing so, they aim to better equip students with the skills and competencies required to successfully navigate their chosen fields of study in an AI-augmented academic environment.

In the context of contemporary education, the purpose of education has shifted from merely acquiring knowledge to developing skills that enable learners to navigate, evaluate, and apply information across diverse cultural and disciplinary landscapes (Bearman et al., 2024). Knowledge is now broadly distributed, and 21st-century learners are expected to engage critically with vast networks of information (Bearman et al., 2024). For EAL students, this means that developing language proficiency alone is insufficient; they must also cultivate academic literacies, intercultural competence, and digital fluency to thrive in contemporary HE.

An additional challenge is that ELICOS EAP programs with direct entry arrangements to universities must comply with the requirements of the Tertiary Education Quality Standards Agency (TEQSA), the education regulator responsible for upholding the quality and reputation of Australian higher education awards. TEQSA stipulates that assessment outcomes must be comparable to other admission criteria for the tertiary course of study. According to the *TEQSA Direct Entry Guide*, these criteria include recognised frameworks such as the Common European Framework of Reference (CEFR) or the Pearson Global Scale of English (GSE), as well as international proficiency tests such as the International English Language Testing System (IELTS) and the Pearson Test of English (PTE) Academic, among others (TEQSA, 2023).

This article explores how EAP programs are responding to these contemporary challenges in HE while retaining compliance with TEQSA requirements. It

examines how curriculum design, assessment practices, and pedagogical approaches within EAP programs are evolving to support the development of academic literacies, intercultural competence, and digital fluency – key capabilities for success in HE. In doing so, it considers the tension between innovation and regulatory accountability, and how programs navigate the dual imperative of preparing students for academic rigour while ensuring assessment comparability with university admission standards. The article begins by examining broader shifts in learning and assessment across the sector before turning to policy and practice at two university language centres - Macquarie University College and Monash College English Language Centre - to illustrate how Australian institutions are adapting to the opportunities and risks posed by global forces, emerging technologies, and GenAI.

THE CHANGING TARGET LANGUAGE USE (TLU) DOMAIN IN HE

Prior to the launch of ChatGPT in November 2022, learning and teaching practices in HE were incrementally adapting to the external forces of changing student demographics and technological advancements. In a study of HE teaching practices, Arkoudis et al. (2013) noted that cultural and linguistic diversity, brought about by internationalisation, is leveraged as a resource for learning and teaching by promoting cross-cultural syndicate groups of students to undertake problem-based, team-based, and project-based learning. Even the traditional teacher-centred lecture has morphed into multimodal events incorporating dialogic, active, and interactive teaching (French & Kennedy, 2016), while emerging pedagogies are focused on engaging students in real-world tasks that furnish them with graduate capabilities such as critical thinking, problem-solving, teamwork, sustainability, and ethical practice (Thomas et al., 2009). In turn, preparatory EAP programs have also been adapting to provide students with opportunities to develop the appropriate socio-cultural, interactional, and discourse competencies necessary for them to participate meaningfully in these educational activities.

A growing trend in HE is micro-modular course design, which breaks traditional curricula into smaller, targeted units. Although common in the USA, UK, and New Zealand, it is a relatively recent development in Australia (Ellis, 2018). This format allows students to tailor their studies to their personal interests, career goals, or skill gaps. For institutions, it streamlines curriculum development, as modules can serve multiple programs. However, the need to assess each module within a

semester raises concerns about fragmented learning and excessive assessment (Harland, 2020), which has been linked to workload and mental health issues among students and staff (Carter & Francis, 2022).

Another key trend in response to increasingly diverse student cohorts is the strategic use of assessment to guide student engagement. With students coming from varied educational and cultural backgrounds, managing cohorts has become challenging; consequently, it is not uncommon for lecturers to use frequent, low-weight summative assessments to encourage regular study habits and improve outcomes. However, this practice has raised concerns that this leads to compartmentalised learning (Harland & Wald, 2020). Additionally, the subsequent proliferation of poorly designed self-marking quizzes has contributed to the overassessment of low-level knowledge and the underassessment of higher-order skills, such as the application of knowledge (Boud, 2020).

The emergence of GenAI provides an opportunity to reset. Since the launch of ChatGPT in November 2022, thought leaders in HE (e.g., Liu & Bates, 2025) are calling for a rethink on the purpose and practice of higher education in the 21st century. They advocate for transformation over incremental change, while retaining the fundamental values of education, through cross-institutional collaboration to develop shared frameworks for governance, pedagogy, and assessment redesign (ibid). One of the most pressing issues is maintaining the integrity of university qualifications. Because GenAI can complete many take-home tasks, stand-alone unsupervised assignments are increasingly vulnerable as sole evidence of unassisted performance.

To support academic integrity efforts, TEQSA's emerging practice toolkit collates strategies and practices from HE institutions across Australia in their management of GenAI and academic integrity issues (TEQSA, 2024). Evidence of the toolkit's three key dimensions is apparent in the emerging trends of learning, teaching, and assessment. The self-assurance mechanisms embedded in the toolkit's dimension of *Process*, for example, can be seen in universities' adoption of strategies, risk management, governance, reporting, and monitoring protocols that support effective practice. The University of Sydney's two-lane approach to assessment, which has been largely influential in Australia, is a good example of this. Lane 1, *assured assessment of learning,* involves secured assessment, such as in-class contemporaneous assessment, viva voces or on-campus examinations, while Lane 2, *assessment as learning*, may incorporate human-AI collaboration (Liu & Bridgeman, 2023). Other universities have similarly leveraged the affordances of

Gen AI to design authentic assessments that target higher-order thinking skills through human-AI collaboration.

It is worth noting that other frameworks have classified assessment into finer levels. One is the AI Assessment Scale by Perkins et al. (2024), which includes five levels, ranging from no AI to AI-assisted idea generation and structuring, AI-assisted editing, AI task completion with human evaluation, and full AI. Another, even more nuanced framework is the six-lane version proposed by Steel (2024), which includes no assistance, simple editing assistance, planning/design assistance, assistance with attribution, GenAI software-based assessment, and 'not applicable,' where AI use is not helpful or relevant to completing an assessment.

The Toolkit's dimension of *People* refers to all levels of stakeholders within the HE ecosystem, such as academic, administrative, support personnel, students, and professional accreditation bodies. Professional development initiatives for staff and coaching initiatives for students in the use of GenAI are evidence of this dimension in action. An illustrative example is the emphasis on cultivating students' evaluative judgement regarding the outputs and processes of generative AI. Anticipating a future where the review of AI outputs is routine, Bearman et al. (2024) advocate emphasising critical thinking and evaluative judgement skills that will help students to evaluate and assess the reliability, relevance, and ethical implications of GenAI content. Cultivating students' capacity for responsible and informed engagement with emerging technologies is seen as key (Bearman et al., 2024; Liu & Bates, 2025).

An example of the toolkit's dimension of *Practice,* which operates at the level of teaching, learning, and assessment, can be found in the increasing use of viva voces and interactive oral assessment in HE. These methods allow instructors to directly engage with students' understanding, making it more challenging to rely on AI-generated responses (O'Riordan et al., 2025; Sotiriadou, 2020; Ward et al., 2023). Moreover, it is argued that oral assessments foster deeper learning by encouraging students to articulate their reasoning and reflect critically in real time (Shaeri et al., 2021; Ward et al., 2023).

CASE STUDIES

Policy and practices at Macquarie University and Monash University show alignment and divergence in their approaches to GenAI. Macquarie University's 2025 draft position statement to GenAI, for example, outlines four core principles:

(1) prioritising purpose over restriction by designing assessments that reflect AI's role in modern work; (2) promoting an emphasis on critical thinking, judgement, and appropriate tool usage over memory recall; (3) encouraging adaptability in using technology for problem solving; and (4) promoting ethical, values-based application of tools. All aspects of curriculum design, including assessment, are expected to adopt a programmatic approach, which entails shifting away from multiple micro-assessment tasks at the unit level to considering how assessment functions across the entire degree. Throughout the first part of 2025, in place of smaller, continuous assessment tasks, which were common in unit design, Macquarie has set a maximum of three assessment tasks for any unit. Plans are also underway to consolidate the number of assessment types to a more defined range within the university's curriculum management system. This work is a part of a well-being response to feedback about the impact of high assessment loads on students and staff.

By contrast, Monash University adopts programmatic assessment approaches, as recommended in TEQSA's guidance (Lodge et al., 2023), within a scaffolded suite that strategically integrates AI and utilises secure assessments where appropriate. In January 2025, it began a two-year Programmatic Assessment and Artificial Intelligence Review project to examine and recalibrate all coursework programs (Monash University, 2025b). This work sits within a broader effort to advance assessment by integrating AI strategically and using secure assessments where appropriate, within a coherent, scaffolded suite of activities that prepares students to operate effectively in the AI era. The principles of programmatic assessment rest on three pillars (Monash University, 2025a). First, feedback is continuous and meaningful, building learner agency and self-regulation. Second, a coherent mix of individual and collaborative assessment methods is selected with a clear educational rationale within the wider program and used holistically over time to provide multiple, scaffolded opportunities to demonstrate achievement across the program. Third, assessment decisions are equitable and credible, triangulating evidence across tasks and course learning outcomes within transparent policy frameworks. Collectively, these principles treat every assessment artefact as evidence, provide learners with multiple opportunities to review, reflect, and integrate feedback, and ensure defensible program-level judgments (Monash University, 2025a). The project is currently underway and is expected to lead to transformative assessment practices, regimes, and tasks across disciplines.

INNOVATIONS AND CHALLENGES IN EAP ASSESSMENT DESIGN

In response to the sweeping changes in the TLU domain, Australian EAP has evolved. EAP programs have traditionally focused on genre-based instruction with academic genres that have highly definable and teachable structures, such as research articles and academic presentations. Genre analysis, such as Swales' (1990) CARS model and Hyland's (2009) work on oral presentations, provides frameworks for understanding discourse structure, signposting, and sociocultural norms. These approaches expose students to target linguistic features and formulaic language, which they can apply in their own discourse. Other common academic genres include literature reviews, annotated bibliographies, case studies, and reports, which promote authentic assessment and transferable academic skills.

Even before the rise of GenAI in late 2022, tools like Grammarly had already raised concerns about academic integrity in EAP programs. With no policies in Australian universities or ELICOS centres addressing them, administrators faced uncertainty about how to classify their use and whether to ban such tools, adapt rubrics, or replace take-home assignments with invigilated assessments (Dinneen, 2021). With the recent and rapid advancement of GenAI, however, questions around how to assess students' language use with reliable and authentic assessment are even more pronounced.

The table below illustrates a range of GenAI tools that can be used to complete typical EAP assessment tasks, often requiring minimal language input from the student. While the GenAI output may include hallucinated sources and fabricated bibliographic citations (Walters & Wilder, 2023) or show weak domain-specific reasoning, they may well pass the assessment, as has been seen in law (Choi et al., 2022), medicine (Panthier & Gatinel, 2023), or humanities and social sciences (Farazouli et al., 2023).

Table 1. GenAI disruption of traditional EAP tasks

Assessment Task	**Illustrative GenAI Tools**
Literature reviews	Paperguide, Elicit, Semantic Scholar
Annotated bibliographies	Writefull, Zotero, ChatGPT
Presentations	CoPilot, Gamma.app, Canva
Writing reports and essays	ChatGPT, CoPilot, Perplexity, Claude, Gemini, Jasper
Videos	Synthesia, Animaker Deck, Galaxy AI, Veed.io, Akool

The realisation that these tasks, in their traditional forms, may have become obsolete as a means of assessing students' language skills is slowly spreading, although there are varying responses among ELICOS centres. At the 2025 University English Centres of Australia (UECA) Assessment Symposium, several centres reported using assignment instructions that explicitly permit or prohibit the use of GenAI, accompanied by rating rubrics that penalise language inconsistent with students' in-class performance. This approach, however, is difficult to police effectively, and it overlooks the fact that commonly used software, such as Microsoft Word, has built-in spelling and grammar checks that can legitimately enhance student output. Making decisions about which tools are allowed or disallowed and verifying compliance can lead to inconsistency among raters and ultimately undermine the reliability of assessment outcomes. It also raises the question of authenticity: if a student can legitimately use GenAI tools in practice in their university degrees, is it reasonable to deny them access or training for the use of these tools while they are in their preparatory program?

Macquarie University College (MQC) employs a sustainable assessment design (Boud & Soler, 2015), which meets the evaluative needs of the EAP program as well as the future needs of students at the university. Course design aims to strike a balance between advancing students' language proficiency while introducing AI literacy and fostering the development of appropriate sociocultural, interactional, and discourse competencies for university study. GenAI collaboration is integrated into learning and teaching activities for drafting, feedback, and comparative evaluation. It is also used in minor summative assessment tasks. Due to the high-stakes nature of the direct entry programs, MQC has allocated most of the assessment as *assured learning,* excluding the use of GenAI, so that students' real language use can be assessed. The table below shows the weighting of GenAI integrated assessments versus non-GenAI tasks, which are invigilated and handwritten.

Table 2. MQC Direct Entry 10 assessment weighting

MQC Direct Entry 10	**Weighting**	**Assessment Task & Weighting**
Assessment that incorporates GenAI	30%	Multimodal reflective journal Researched Oral Presentation
Assessment that excludes GenAI	70%	Integrated reading to write task Integrated listening to speak task Integrated reading, listening to write task

As an interim measure when GenAI first emerged, MQC replaced the traditional research essay with a process-based task focused on drafting, deadlines, and feedback. Submissions were supported by a resource folder containing original materials cited in student essays, along with follow-up interviews to ensure academic integrity. Although beneficial for developing academic skills, this approach was highly time-consuming and offered limited assessment of students' independent language use.

Concurrently, MQC was transitioning from separate skills tests to integrated skills assessments, which are better aligned with the TLU domain, where students draw on source texts to produce written or spoken output, such as essays and verbal reports. The use of these tasks revealed students' need for stronger summary and paraphrasing skills. Consequently, the research essay was removed to make space for coursework that prioritises targeted language development, feedback literacy, and multiple feedback points through self, peer, AI, and teacher review.

Monash University English Language Centre (MUELC), on the other hand, has adopted a dual strategy that combines assured assessment with guided, AI-assisted assessment. For assured assessments across all modules, the large, end-of-course online examinations introduced during the COVID-19 period (and proctored over Zoom) have been replaced by regular, lower-stakes online reading, listening, and writing tests delivered in Moodle using a secure browser called Safe Exam Browser and invigilated in class by teachers. Speaking is assessed directly through a range of tasks, including digital storytelling, video submissions, spoken reflections, group speaking tasks, tutorial discussions, and oral presentations, which provide multiple samples of students' performance.

For extended writing tasks, such as the Research Task in Module 4 (formerly the Monash English Bridging to university course), MUELC embeds GenAI authentically as part of the research process, while preserving the requirement

that students produce the final paper independently. Across the teaching weeks, students receive explicit instruction on using GenAI to brainstorm, develop structure and organisation, locate sources, and generate provisional reference lists. They are also taught to critically evaluate GenAI outputs, from checking for factual errors to assessing superficiality, cultural stereotyping, and fabricated sources. Students then learn how to substantiate their claims with appropriate academic literature rather than relying on AI-generated sources. Final papers are written without assistance and are assessed by the teacher, supported by a one-to-one interview in which students explain their ideas, demonstrate conceptual and process understanding, and discuss their use of GenAI.

Following a small-scale pilot involving 180 students and eight teachers (Hellema et al., 2024), MUELC introduced the AI-incorporated Research Task in the Module 4 program. In 2025, the task was refined to require each student to draft in a version-tracked online document (with edit history) and to maintain a research folder containing all source texts; both artefacts are accessible to teachers. While no mechanism can guarantee wholly unassisted writing, combining process history with a 10-minute viva voce provides teachers with deep insight into students' writing process, research practices, conceptual understanding, and transparency around GenAI use. Students reported feeling better equipped to collaborate with GenAI in a productive and ethical manner. MUELC reported a reduction of over 80% in suspected academic integrity breaches in 2025 (Sadeghpour, Module 4 Program Leader, 2025, personal communication), freeing staff time for targeted teaching and student support. In Module 4, the Research Task is paired with an invigilated Reflective Writing test. Together, the two tasks provide triangulated evidence of students' writing ability. The model is currently under consideration for adoption in lower-level modules (Modules 2 and 3), where GenAI is already used to support the creation of multimodal learning artefacts.

GENAI PRODUCTIVITY GAINS IN EAP CONTEXTS

The literature highlights growing productivity gains from GenAI in language teaching and learning (Chapelle, 2025; Koraishi, 2023; Law, 2024). In Australian EAP programs, practitioners report that GenAI especially boosts efficiency in routine tasks, ideation, and content generation. Importantly, they stress the need for human-led critical evaluation, editing, and review of language assessment where construct-relevant decisions remain essential (e.g., Starford, 2024; Tran et al., 2024). In practice, this means EAP practitioners use GenAI to expand and

update assessment materials and to draft feedback, but do so within precise specifications of what constitutes evidence of students' language proficiency in assured assessment and how AI use is designed and declared in open assessment. This mirrors recent sector advice to balance assured and open assessments while preserving academic integrity safeguards (Liu & Bridgeman, 2023; Lodge et al., 2023; TEQSA, 2024).

GenAI can accelerate EAP content and assessment development by generating banks of reading and listening texts, writing task prompts, and speaking scenarios based on university-relevant themes. By specifying genre features and targeted language level in the prompt, teachers can tailor GenAI outputs to diverse curriculum needs. This streamlining revolutionises resource-heavy assessment development, supporting frequent task rotation and assessment security, and helps keep content current with disciplinary themes common in Australian universities. However, thorough human vetting is always required to ensure outputs are accurate, level-appropriate, and free of inadvertent plagiarism or copyright risk (Starford, 2024; Tran et al., 2024).

Another area of high impact is GenAI's enhancement of feedback processes. Teachers can use GenAI to draft formative comments that flag surface-level features and map them to rubric indicators. This automated generation of feedback frees teachers to focus on providing more in-depth feedback and feedforward. When students use GenAI independently to seek feedback, it can provide useful responses that are timely, easily accessible, and presented in a clear, digestible language that supports understanding (Henderson et al., 2025). GenAI can also help students identify areas for improvement in their work. Nevertheless, in skill-based domains such as writing, deeper learning and skill development can be undermined if students lack effective prompting skills, leading them to receive direct edits or suggested changes rather than formative feedback that prompts reflection on errors and scaffolds the reorganisation or rephrasing of their ideas (Luhach, 2024).

Interestingly, empirical research comparing students' perceptions of AI-generated and teacher feedback reveals that while students value the immediacy and availability of AI responses, they place greater trust in teacher feedback, describing it as more "relational, contextualised, relevant, and reliable" (Henderson et al., 2025, p. 12). This suggests a complementary nature between the two sources of feedback, in which AI handles first-pass feedback, while teachers provide more tailored feedback, calibration, and next-step guidance (Henderson et al., 2025).

This aligns with Australian feedback literature that emphasises learner uptake and interaction over mere information delivery (Jensen et al., 2023). When turnaround times are reduced or resources are limited, teachers can combine AI-drafted feedforward with self-assessment using rubrics, exemplars, and brief viva checks to preserve the social aspect of feedback (Boud & Molloy, 2013; Jensen et al., 2023).

Scalability gains are particularly notable in ELICOS centres with multiple intakes across 10- to 20-week programs. GenAI can assist teams in producing level-specific versions of reading-into-writing tasks, alternative prompts for discussion questions, or refreshed teaching materials that foreground current Australian debates (e.g., AI in research and academic integrity). Where direct entry comparability is required, productivity gains can be channelled into developing learning and practice materials (e.g., practice tasks or annotated exemplars) while secured, invigilated assessments continue to elicit unassisted language production for decision-making, consistent with TEQSA's aforementioned ELICOS Direct Entry expectations (TEQSA, 2023).

While GenAI can improve efficiency, it does not automatically produce classroom-ready materials or assessment tasks. The initial exploration by Chapelle et al. (2024) of the use of ChatGPT in generating reading texts shows that early versions of AI-generated texts were longer than required, included inaccurate information, and used vocabulary more advanced than specified; consequently, producing fit-for-purpose texts required multiple iterations, more detailed prompting, and significant post-editing. In assessment development, GenAI may generate workable multiple-choice stems and answer keys, as human assessment developers do; however, it often struggles to create plausible distractors (Chun & Barley, 2024). Uncritical use of GenAI outputs can thus yield teaching and assessment content that contains factual errors, perpetuates cultural biases, or is pitched at an inappropriate level and therefore unfair (Chapelle et al., 2024). Students may also find AI feedback repetitive or less trustworthy when it is not tailored to their work and to the program's materials, assessment tasks, and rubrics.

To mitigate these risks, EAP programs should implement sound assessment and feedback design principles, require thorough review and vetting processes, and build staff capability. Providers should raise teachers' and assessment developers' critical AI literacy (Chapelle, 2025) and strengthen their evaluative judgement capability (Bearman et al., 2024). One practical step is to require staff to critique AI outputs against robust criteria and annotated exemplars before adapting and

integrating these outputs into assessment materials. Established item-writing guidelines, such as those outlined by Haladyna (2004) for multiple-choice items, remain useful supports in this process.

IMPLICATIONS FOR ENGLISH LANGUAGE TEACHING AND ASSESSMENT

Beyond improving efficiency, GenAI has broad implications for English language teaching and assessment due to its cognitive impact on users. Recent research on the thought processes of GenAI-assisted essay writing reports that GenAI support reduces engagement of the brain networks involved in internal idea generation and monitoring. It is also associated with weaker memory and a diminished sense of ownership (e.g., poorer ability to quote one's own text) compared with unaided ('brain-only') writing and search-engine-assisted writing (Kosmyna et al., 2025). The authors characterise this as 'cognitive debt': short-term fluency gained with GenAI at the expense of deeper learning, with carry-over effects when students later write without assistance (Kosmyna et al., 2025). These insights have implications for both language teaching and assessment. The authors recommend instruction that engages unaided and search-engine-assisted writing skills, as well as GenAI-supported writing, to develop a broad range of skill sets.

As EAP programs in Australia continue to operate in the age of GenAI, the question is how to integrate these tools in ways that preserve the existential function of developing learners' language proficiency and academic literacies for university studies while developing their AI literacies and ethics.

Curricula should specify where GenAI is permitted in language learning and assessment, as well as how it is restricted in assured assessment. The two-lane approach to assessment (Liu & Bridgeman, 2023), alongside more detailed frameworks such as Perkins et al.'s (2024) five-level AI Assessment scale and Steel's (2024) six-lane framework, offers a pragmatic path: a strategic combination of open, learning-oriented tasks in which students use GenAI transparently for brainstorming, sourcing, and drafting support under teacher mediation and secured, invigilated checkpoints that elicit unassisted performance for progression or university entry decisions.

Given the reported reduction in neural engagement with the use of GenAI in writing (Kosmyna et al., 2025), programs could consider withholding the introduction of GenAI tools in early stages. They could start with brain-only writing activities to develop students' ability to produce English independently,

before progressively incorporating search-engine- and GenAI-supported activities. Assured checkpoints should then collect a sample of unaided planning and information integration (e.g., reading-to-write responses), while open tasks make the use of GenAI visible and teachable.

Pedagogically, genre-based teaching may remain central, but it should be extended to include critical work on prompts, acknowledgement and citation of AI-mediated ideas, verification of claims, and comparison of human- and AI-produced exemplars. To counter 'cognitive debt' (Kosmyna et al., 2025), teachers can embed practices that activate thinking and planning in the writing process. Where AI assistance is used, they can design tasks that require students to explain their evaluative judgments and present their critical reflections. Pairing process evidence, such as planning notes, version histories, and records of incremental improvement based on feedback, with brief viva voce sessions should be normalised so that teachers can triangulate learning without defaulting to policing or surveillance.

Professional development for assessment developers and teachers must now include AI literacy alongside language assessment literacy. These staff require access to institutionally approved AI tools and opportunities to develop practical AI use capabilities, with an understanding of tool affordances and limitations, so they can generate outputs supported by reliable sources and calibrated to a specified level of linguistic sophistication. They also need to develop skills in auditing outputs for level, accuracy, and bias; designing AI-assisted feedback that students can utilise; and documenting AI use in a manner consistent with institutional policy and sector standards. Equally important are privacy and data security protocols for handling student work. Time saved by AI-enabled drafting and first-pass feedback should be deliberately reallocated to high-value activities, such as diagnostic assessment, oral language development, modelling of paraphrasing, summarising, citation, and GenAI use, and targeted language enhancement activities, rather than being absorbed by administrative tasks. Program leaders can support this shift by providing shared prompt libraries or rubric-aligned comment banks.

These changes reimagine the role of the EAP practitioner. Beyond serving as a materials writer and assessor, the practitioner becomes a curator, a coach, and a designer of evidence: curating task inputs to sustain local authenticity; coaching evaluative judgement so students can appraise AI outputs against criteria; and designing coherent, and, where feasible, programmatic assessment that combines

AI-supported learning with credible, comparable assessment decisions. The practitioner also acts as a policy translator between institutional governance and classroom practice, ensuring that AI use declarations, authorship checks, and assured assessment methods are proportionate and educative. When used deliberately, GenAI can augment teacher expertise in areas such as expanding variation and currency in tasks and accelerating feedback cycles while allowing EAP teachers to focus on the human work of making expert judgements and providing targeted guidance and support that enhances student learning.

RECOMMENDATIONS FOR CLASSROOM RESEARCH

The integration of GenAI into academic and professional domains has fundamentally altered how language is produced, refined, and evaluated by educators, students, and employers, prompting new standards for communication, critical thinking and authorship across diverse contexts. As such, it is reasonable for English language development to assume a degree of human-AI collaboration. The question remains, however, as to what level of human-AI collaboration is appropriate, for which tasks, and under what conditions in EAP programs that assess a student's English language proficiency. There is certainly an argument to be made that, as English language use evolves, so too must the ways of assessing EAL students' readiness for university study.

For EAP program providers, classroom-based validation and external benchmarking are necessary to establish evidence that new assessment regimes, which include AI-incorporated assessments, yield valid, fair, and comparable inferences about proficiency, comparable to those of international language proficiency tests. A related question is the continuing role of international proficiency tests in determining readiness, particularly where students' skilful and ethical use of GenAI may mediate their ultimate academic success. Research should therefore examine how unaided performance and AI-mediated performance jointly inform decisions on progression and entry.

Nonetheless, to critically engage with, refine, and ethically deploy AI-generated content within academic contexts, students must possess a strong level of language proficiency. It remains the task of EAP programs not only to develop core linguistic skills, but also to cultivate a broader set of academic and digital literacies. These include critical reading, information literacy, genre awareness, and socio-pragmatic competence alongside digital fluency and ethical reasoning. By fostering these skills, EAP programs empower students to collaborate effectively with

teachers, peers, and GenAI, navigate complex academic tasks, and contribute confidently to university-level discourse.

EAP educators are uniquely positioned to observe and evaluate how students interact with GenAI in academic tasks, providing valuable insights into the linguistic competencies required to use these tools effectively and informing the pedagogical shifts needed to support ethical and critical engagement. Several interesting research pieces provide direction for future research options in EAP, for example:

1. Hawkins, Taylor-Griffiths and Lodge (2025) explore how students use GenAI for feedback during essay writing, finding that feedback literacy and the ability to evaluate essay outputs are significant predictors of essay performance.

2. Haidar and Tassis (2025) investigate the perceptions and integration of GenAI among EAP instructors, finding that rather than seeing GenAI as transformative, instructors largely view it as a supplementary tool to enhance existing practices.

3. Roe, Perkins and Tregubova (2024) introduce the EAP AI Assessment scale to guide ethical and pedagogically sound integration of GenAI into EAP instruction. Their five-level scale has use cases at each level offering recommendations for implementations and future research in this field.

CONCLUSION

EAP programs in Australian higher education are undergoing significant transformation as they respond to the dual pressures of technological innovation and regulatory compliance. Through evolving curriculum design, assessment reform, and pedagogical adaptation, these programs aim to equip students with the academic literacies, intercultural awareness, and digital competencies necessary for success in university studies. The case studies of Macquarie University College and Monash College English Language Centre illustrate how institutional policy and classroom practice are beginning to engage with the possibilities and challenges introduced by emerging technologies such as GenAI. Ultimately, the capacity of EAP programs to strike a balance between 'innovation' and 'accountability' will determine their effectiveness in preparing students for the demands of contemporary higher education. In practice, this balance, as illustrated

in the two centres' examples, can entail pairing transparent, GenAI-supported learning tasks with secure, invigilated checkpoints that evidence unaided performance; documenting process evidence (e.g., drafts, prompts, vivas); and building staff capability in AI literacy and evaluative judgement**.** Ensuring responsible and effective integration will require a collaborative, interdisciplinary approach that brings together educators, technology specialists, applied linguists, and policymakers to co-design frameworks that uphold academic integrity while enabling innovation and supporting comparability with external admissions criteria.

AUTHORS

Dr Phuong Tran is the Manager for Assessment at Monash College. She specialises in educational assessment and measurement, language teaching and teacher education. Her current interests are transition education and quality assessment design in the face of generative artificial intelligence.

Cara Dinneen is Education Manager for English Language Programs and English Medium Instruction at Macquarie University College. Cara is also the convenor of the English Australia Assessment Special Interest Group, and her professional interests include learning, teaching, and assessment design for high-stakes EAP programs.

REFERENCES

Arkoudis, S., Watty, K., Baik, C., Yu, X., Borland, H., Chang, S., . . . Pearce, A. (2013). *Finding common ground: Enhancing interaction between domestic and international students in higher education.* https://doi.org/10.1080/13562517.2012.719156/

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, *49*(6), 893–905.

Boud, D. (2020). Challenges in reforming higher education assessment: a perspective from afar. *RELIEVE*, *26*(1), art. M3. https://doi.org/10.7203/relieve.26.1.17088

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education, 38*(6), 698–712. https://doi.org/10.1080/02602938.2012.691462

Boud, D., & Soler, R. (2015). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, *41*(3), 400–413. https://doi.org/10.1080/02602938.2015.1018133

Carter, M. A. & Francis, A.P. (2022). Introduction. In Francis, A.P. & Carter, M. A. (Eds.), *Mental health and higher education in Australia* (pp. 3–13). Springer.

Chapelle, C. A. (2025). Generative AI as game changer: Implications for language education. *System*, *132*(2025), 103672. https://doi.org/10.1016/j.system.2025.103672

Chapelle, C.A., Baghestani, S. Beck, J., Kurt, S., Kusumaningrum, W., Lim, H. J., & Yin, S. (2024). Exploring Open AI in language teacher professional development. In C. A. Chapelle, G. H. Beckett, & J. Ranalli (Eds.), *Exploring artificial intelligence in applied linguistics* (pp. 217–242). Iowa State University Digital Press. https://doi.org/10.31274/isudp.2024.154.13

Choi, J. H., Hickman, K.E, Monahan, A.B., & Schwarcz, D. (2022). ChatGPT goes to law school. *Journal of Legal Education*, *71*(3), 387–400.

Chun, J. Y., & Barley, N. (2024). A Comparative analysis of multiple-choice questions: ChatGPT-generated items vs. human-developed items. In C. A. Chapelle, G. H. Beckett, & J. Ranalli (Eds.), *Exploring artificial intelligence in applied linguistics* (pp. 118–136). Iowa State University Digital Press. https://doi.org/10.31274/isudp.2024.154.08

Dinneen, C. (2021). Students' use of digital translation and paraphrasing tools in written assignments on direct entry English programs. *English Australia Journal, 37*(1), 40-49. https://eajournal.partica.online/digital/english-australia-journal-371/flipbook/48/

Ellis, C. (2018). The practicalities of using assessment management to develop evaluative judgement. In D. Boud, R. Ajjawi, P. Dawson, & J. Tai (Eds.), *Developing evaluative judgement in higher education: Assessment for knowing and producing quality work* (pp. 127–135). Taylor & Francis Group.

Farazouli, A., Cerratto-Pargman, T., Bolander-Laksov, K., & McGrath, C. (2023). Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices. *Assessment & Evaluation in Higher Education*, *49*(3), 363–375. https://doi.org/10.1080/02602938.2023.2241676

French, S., & Kennedy, G. (2016). Reassessing the value of university lectures. *Teaching in Higher Education*, *22*(6), 639–654. https://doi.org/10.1080/13562517.2016.1273213

Haidar, R., & Tassis, A. (2025). *Power, pedagogy, and algorithms: An exploratory study of EAP instructors' critical engagement with Generative AI.* https://doi.org/10.21203/rs.3.rs-7014667/v1

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ed.).
Routledge.

Harland, T. (2020). *University challenge: Critical issues for teaching and learning*: Routledge.

Harland, T., & Wald, N. (2020). The assessment arms race and the evolution of a university's assessment practices. *Assessment and Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2020.1745753

Hawkins, B., Taylor-Griffiths, D., & Lodge, J. M. (2025). Summarise, elaborate, try again: exploring the effect of feedback literacy on AI-enhanced essay writing. *Assessment & Evaluation in Higher Education*, 1–13. https://doi.org/10.1080/02602938.2025.2492070

Hellema, R., Sadeghpour, M. & James, A. (2024). Preserving fidelity: Adapting academic writing assessment in the age of GenAI. *Monash College Inquiry-Driven Education in Action (IDEA) Program: Research Projects 2024*, pp. 62–71.

Henderson, M., Bearman, M., Chung, J., Fawns, T., Buckingham Shum, S., Matthews, K. E., & de Mello Heredia, J. (2025). Comparing generative AI and teacher feedback: Student perceptions of usefulness and trustworthiness. *Assessment & Evaluation in Higher Education*, 1–16. https://doi.org/10.1080/02602938.2025.2502582

Hyland, K. (2009). *Academic discourse*. Bloomsbury Publishing.

Jensen, L. X., Bearman, M., & Boud, D. (2023). Feedback encounters: towards a framework for analysing and understanding feedback processes. *Assessment & Evaluation in Higher Education*, *48*(1), 121–134. https://doi.org/10.1080/02602938.2022.2059446

Koraishi, O. (2023). Teaching English in the age of AI: Embracing ChatGPT to optimize EFL materials and assessment. Language Education Technology, *3*(1), 55–72. https://langedutech.com/letjournal/index.php/let/article/download/48/37

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X. H., Beresnitzky, A. V., et al. (2025). *Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task*. Cornell University. https://doi.org/10.48550/arXiv.2506.08872

Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open*, *6*(2024) 100174. https://doi.org/10.1016/j.caeo.2024.100174

Liu, D. Y. T., & Bates, S. (2025, January). *Generative AI in higher education: Current practices and ways forward* [White paper]. Association of Pacific Rim Universities. https://biblioteca.unisced.edu.mz/bitstream/123456789/3658/1/APRU-Generative-AI-in-Higher-Education-Whitepaper_Jan-2025.pdf

Liu, D., & Bridgeman, A. (2023, July 12). *What to do about assessments if we can't out-design or out-run AI?* The University of Sydney. https://educational-innovation.sydney.edu.au/teaching@sydney/what-to-do-about-assessments-if-we-cant-out-design-or-out-run-ai/

Lodge, J. M., Howard, S., Bearman, M., Dawson, P., & Associates. (2023). *Assessment reform for the age of Artificial Intelligence.* Tertiary Education Quality and Standards Agency. https://www.teqsa.gov.au/sites/default/files/2023-09/assessment-reform-age-artificial-intelligence-discussion-paper.pdf

Luhach, S. (2024). An exploratory study of AI assisted feedback and formative assessment of writing skills. *2024 International Conference on Artificial*

Intelligence and Emerging Technology (Global AI Summit), 1277–1281. https://doi.org/10.1109/GlobalAISummit62156.2024.10947841

Monash University. (2025a). *Programmatic approaches to assessment*. https://www.monash.edu/learning-teaching/TeachHQ/Assessment/PAAIR/programmatic-approaches-to-assessment

Monash University. (2025b). *Programmatic assessment and AI review (PAAIR)*. https://www.monash.edu/learning-teaching/TeachHQ/Assessment/PAAIR

O'Riordan, F., Thangaraj, J., Girme, P. & Ward, M. (2025). Interactive oral assessment: Staff perceptions, challenges and benefits of this robust, authentic assessment design approach. *Innovations in Education and Teaching International*. https://doi.org/10.1080/14703297.2025.2477160

Panthier, C., & Gatinel, D. (2023). Success of ChatGPT, an AI language model, in taking the French language version of the European Board of Ophthalmology examination: A novel approach to medical knowledge assessment. *Journal Français d'Ophtalmologie*, 46(7), 706–711. Scopus. https://doi.org/10.1016/j.jfo.2023.05.006

Perkins, M., Furze, L., Roe, J., MacVaugh, J. (2024). The artificial intelligence assessment scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice*, 21(6). https://doi.org/10.53761/q3azde36

Thomas, T., Petocz, P., Rigby, B., Clark-Murphy, M., Daly, A., Dixon, P., et al. (2009, November). Embedding generic skills means assessing generic skills. In *ATN Assessment Conference 2009: Assessment in Different Dimensions* (p. 321).

Roe, J., Perkins, M., & Tregubova, Y. (2024). *The EAP-AIAS: Adapting the AI Assessment Scale for English for Academic Purposes. https://arxiv.org/pdf/2408.01075*

Shaeri, S., Logan, D., & Krautloher, A. (2021). Evaluating competency development using interactive oral assessments. *Research in Engineering Education Network and the Australasian Association for Engineering Education Conference* (pp. 1–10). Engineers Australia, Perth, Western Australia.

Sotiriadou, P., Logan, D., Daly, A., & Guest, R. (2020). The role of authentic assessment to preserve academic integrity and promote skill development and employability. *Studies in Higher Education*, *45*(11), 2132–2148. https://doi.org/10.1080/03075079.2019.1582015

Starford, A. (2024). *Utilising generative AI for ELICOS resource development*. [Conference presentation]. English Australia Conference. https://www.englishaustralia.com.au/documents/item/2375

Steel, A. (2024). *2 lanes or 6 lanes? It depends on what you are driving: Use of AI in assessment*. https://www.education.unsw.edu.au/news-events/news/two-six-lanes-ai-assessment

Swales, J. M. (1990). *Genre analysis*. Cambridge University Press.

Tertiary Education Quality Standards Agency. (2023). *ELICOS direct entry guide.* https://www.teqsa.gov.au/sites/default/files/2023-11/ELICOS-Direct-Entry-guide.pdf

Tertiary Education Quality Standards Agency. (2024). *Gen AI strategies for Australian higher education: Emerging practice.* https://www.teqsa.gov.au/guides-resources/resources/corporate-publications/gen-ai-strategies-australian-higher-education-emerging-practice

Tran, P., Rooney, M., & Bastock, J. (2024, August 31). *Enhancing language assessment security through teacher training and collaborative design*. [Symposium presentation]. UECA Assessment Symposium, Melbourne, Australia. https://online.ueca.edu.au/presentations/presentation-phuong-tran-mark-rooney/

Walters, W.H. & Wilder, E.I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports*, *13,* 14045. https://doi.org/10.1038/s41598-023-41032-5

Ward, M., O'Riordan, F., Logan-Fleming, D., Cooke, D., Concannon-Gibney, T., Efthymiou, M., & Watkins, N. (2023). Interactive oral assessment case studies: An innovative, academically rigorous, authentic assessment approach. *Innovations in Education and Teaching International*, *61*(5), 930–947. https://www.tandfonline.com/doi/full/10.1080/14703297.2023.2251967

The ABCs of Psychometric Validity: A Practical Guide for English Language Teachers

Talia Isaacs[1]

UCL Institute of Education, University College London

ABSTRACT

This article offers an accessible discussion of assessment literacy and validity in the context of language testing and assessment practice. Assessment literacy can be broadly defined as the skills and knowledge teachers have about the functions and roles of assessment in educational contexts. It is also related to the ability to choose various forms of assessment to serve an intended purpose, as well as an awareness of critical assessment-related issues, such as validity, fairness, ethics, and the effects of assessment on students and other stakeholders. This article relates assessment literacy to the concept of psychometric validity, which can help English language teachers evaluate their own and others' assessment practices through explicit and practical guides. Validity is a complex concept that requires clear explanations and examples tailored for teachers. Through this article, readers can enhance their understanding of core assessment concepts and learn how to implement positive assessment practices.

[1] Correspondence: Talia Isaacs (talia.isaacs@ucl.ac.uk); https://orcid.org/0000-0003-4302-3379

ISSN 1834-4712 (Online).

DOI: 10.5281/zenodo.17590719

Citation: Isaacs, T. (2025). The ABCs of psychometric validity: A practical guide for English language teachers. *University of Sydney Journal of TESOL*, *4*, 137–162. https://doi.org/10.5281/zenodo.17590719

Keywords: assessment literacy, classroom-based assessment, educational assessment, language assessment, open science, psychometrics, teacher education, validity, washback

INTRODUCTION

National educational reform is a global phenomenon. One of governments' national strategic priorities is economic competitiveness with other countries. This necessitates a highly educated, skilled workforce, inevitably leading to periodic educational reform (OECD, 2021). Decisions on education reform by governments around the world are often based on statistical records of students' improved attainment levels derived from standardised assessments, placing assessment at centre stage.

Governments tend to emphasise national standards-based assessment over teacher assessment, which is often viewed sceptically in terms of quality and consistency (Leung & Rea-Dickins, 2007). Concerns, particularly in debates about teacher assessments supplanting standardised testing, include but are not limited to potential bias—for example, in relation to gender or socioeconomic status (Burgess, 2015), claims of grade inflation (Guardian, 2021) and disagreement among teachers regarding classroom assessment practices (Lopez & Pasquini, 2017). While variations exist across countries, governments frequently overlook the localised technical and pedagogical challenges faced by teachers. Many language teachers lack formal training in assessment design. Katz (2012) rightly points out that for many teachers, "assessment is often treated as an add-on to the primary focus of instruction, and as a source of information about the outcome of carefully planned classroom activities," and their "decision as to whether or not to link instruction and assessment has already been made for them" (p. 69).

A new and pressing challenge in this landscape is the rise of Generative AI (GenAI), which is rapidly transforming the way language is produced, evaluated, and understood. GenAI tools (e.g., ChatGPT, Copilot, etc.) can generate fluent, coherent texts that mimic human writing, raising concerns about the validity of traditional language assessments. These tools blur the boundaries between original student output and AI-assisted writing, making it increasingly difficult to assess authentic language ability. As a result, educators and policymakers must reconsider assessment design, integrity protocols, and the very constructs being measured in language education (Harding, 2026). The integration of GenAI into educational contexts necessitates a reevaluation of assessment literacy, particularly among

teachers who may lack the necessary skills to navigate these emerging complexities (Cui et al., 2025).

In light of the ongoing education reforms in various national contexts worldwide, serious attention should be given to supporting English language teachers in improving their assessment practices (Harding & Kremmel, 2017; Inbar-Laurie, 2012; Taylor, 2009). This article aims to promote teachers' assessment literacy by describing and explaining the ABCs of validity in assessment with practical guidelines. The article addresses the issues of classroom assessment, introduces the concept of assessment literacy, and provides an overview of language assessment types in education. It then differentiates between psychometric and edumetric assessments before introducing the notion of validity. The article provides critical guidelines for enhancing validity in classroom assessment. We particularly unpack "psychometric validity", which has been highly influential in the fields of educational and language assessment.

TEACHER ASSESSMENT IN THE LANGUAGE CLASSROOM

For classroom teaching, assessing language ability and learning attainment is an integral part of the language teaching and learning process. Teacher-made assessments for gauging the attainment of student learning outcomes in relation to curricular goals and content coverage should be carefully designed. For English language teachers, many classroom decisions depend on the outcomes of language assessment practices, such as determining what students know and where they need support, either individually or collectively. For students, the ways in which their learning performance has been assessed may well influence their learning attention and motivation (Green, 2018; Poehner & Infante, 2017). In addition to assessments used in the classroom, students may also be required to take externally imposed, often mandatory, high-stakes assessments (e.g., state-wide or national assessments).

Throughout this article, the term 'assessment' is used in a broad sense to refer to the gathering of information about what students know and whether and how they can apply their knowledge (Bachman, 2004). A test is just one high-profile type of assessment that involves quantifying performance quality through the use of a scoring system (e.g., a rating scale or a machine scoring algorithm; Isaacs, 2016). However, there are many other kinds of assessments—that is, ways of systematically gathering information about students' learning through individual or a set of performances (e.g., portfolios, conferencing, presentations, interviews,

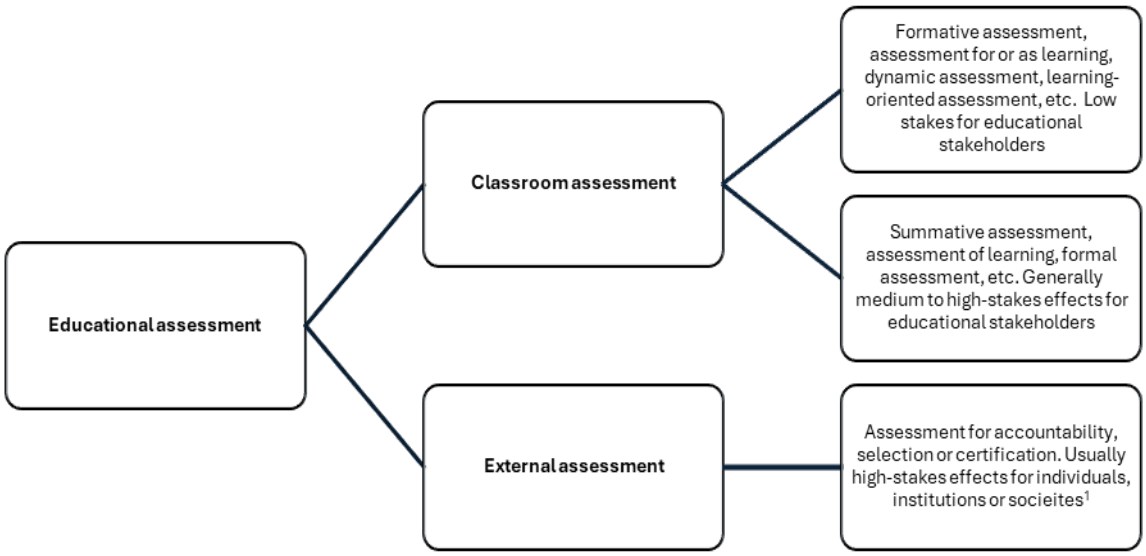
quizzes, etc.). In this article, we use the term 'tests' only when that is the prescribed form of assessment, at the exclusion of other forms of assessment. It is often used for decision-making that has the potential to affect test-takers' lives and may also have repercussions for other educational stakeholders, institutions, and, from a macro-level perspective, society at large.

The *Standards for Educational and Psychological Testing* (hereafter, the *Standards*; AERA, APA & NCME, 2014), which is updated in a new edition every few years, presents an evolving view of psychometric validity that can be viewed as the gold standard of the day; however, it is written for measurement and assessment experts and researchers (Pitts & Naumenko, 2016), so they can use it as a reference point to guide assessment development and use. It is not written for teachers and is not designed to engage teachers with the core concepts of psychometric validity. In education, the concept of validity in relation to assessment holds special significance for teachers to be aware of (Popham, 2017). This reinforces the need to make key psychometric concepts and orthodox contemporary thinking accessible to teachers.

TYPES OF LANGUAGE ASSESSMENT IN EDUCATION

Figure 1 presents a hierarchical structure of educational assessments, including classroom and external assessments. This notwithstanding, for the purpose of our discussion, it may be useful to define these types of assessments separately. However, we do acknowledge they need not *necessarily* be polarised in the way we have portrayed here. First, classroom assessment encompasses formative assessment, which focuses on assessment for and as learning (i.e., process-oriented assessments designed to foster learning), and summative assessment, which focuses on *assessment of learning* (i.e., outcome-oriented assessments).

Figure 1. Educational assessments within and beyond the classroom



[1] Externally imposed assessments may be administered inside or outside of the classroom.

In Figure 1, the word 'stakes' refers to the nature and level of impact or consequences of assessment use on test-takers. Classroom assessment includes a range of assessments:

- *Formative assessment* is used to check whether learning is 'formed' (e.g., through the use of questions, activities, presentations, quizzes, and so forth. Teachers can use this information to inform their teaching. Formative assessment is ongoing throughout the teaching period.

- *Summative assessment* is a summary of student performance, which is taken to be a proxy for learning, at the end of the course. The term' summative assessment' is often used interchangeably with 'assessment of learning' because the purpose of such an assessment is to collect summarised information on student learning performance after the teaching and learning period has been completed. Midterm and final tests are summative assessments and are often known as achievement tests (whether students

have achieved the expected learning outcomes assessed by tests or examinations).

- *Assessment for learning* is often used interchangeably with the term formative assessment. However, it tends to be broader since teachers may not only focus on the learning outcomes that will be assessed. Some of the students' learning challenges or difficulties may not always be directly related to the learning outcomes (e.g., how to use the target language in everyday situations or beyond the test).

- *Assessment as learning* only differs from assessment for learning in that students are at the centre of this approach. It emphasises student autonomy in assessing their own learning as they move through the term. Students regulate and self-assess their own learning, monitor their learning progress and challenges, and work to resolve those challenges. To achieve this, however, teachers need to provide students with guidance on how to regulate and assess their learning. Approaches that integrate assessment as learning include dynamic assessment (see Poehner & Infante, 2017) and learning-oriented assessment (Turner & Purpura, 2016).

Notably, some forms of classroom assessment serve both formative and summative purposes. For example, an essay that counts for marks and receives a formal grade (summative) may be accompanied by feedback through the teacher's comments or reflections (formative) to help the student think more deeply about the topic, foster skill development, or enhance awareness-raising. Although classroom and external assessments, as illustrated in the second level of Figure 1, are sometimes depicted as being in opposition to one another, some integrated models of learning-oriented assessment show that they can operate in tandem to promote learning (Jones & Saville, 2016; Turner & Purpura, 2016) or have blurred boundaries (Fulcher, 2024).

In Figure 1, formative assessments and assessments for learning are often considered low-stakes because their effects on students' lives are low or temporary. Summative assessments are often considered medium- or high-stakes because their effects on students' lives can be impactful and more permanent (e.g., passing or failing a unit of study, grade records in academic transcripts, graduation, college or university admission; see Kunnan, 2018). It should be noted that decisions based on assessment of learning can be used both inside and outside the classroom, with consequences for various stakeholders (not just students). For

example, an institution may use student grades or results from an external standardised test to judge teachers' teaching effectiveness or decide on future school or program funding or surveillance (e.g., Gonzalez & Firestone, 2013).

Finally, Figure 1 illustrates that language teachers are often confronted with external, mandatory, and standardised assessments (e.g., national tests, standards-based assessments, school admission or university or college entrance examinations, commercial language proficiency tests). External, standardised tests are often influenced by measurement theory from the fields of psychometrics and educational measurement. This concerns not only the extent to which a test accurately and reliably assesses the information it claims to assess, but also the way test scores are inferred from learners' performance and faithfully and meaningfully capture learners' underlying language ability, which is the focus of the assessment (Newton & Shah, 2014). Assessment of learning, particularly in the form of external, standardised tests (as opposed to teacher assessments that relate to instructional targets), can sometimes clash with pedagogical goals and lead to a narrowing of the curriculum (e.g., teaching to the test; Phakiti & Isaacs, 2021).

PSYCHOMETRIC APPROACH TO LANGUAGE ASSESSMENT

The psychometric perspective aims to develop generalisable theories of the human mind, including aspects of students' knowledge and ability that are not directly observable but can be inferred from systematic, direct, or indirect observations of behaviour or performance. This perspective has substantially influenced the field of language assessment, with dominant theories of language proficiency centring on linguistic knowledge and the ability to deploy that knowledge (Bachman & Palmer, 2010). To infer proficiency level, assessment stimuli in the form of items or tasks are often used to elicit snapshots of language use behaviours, which, in turn, are used to infer underlying language abilities or other associated constructs of interest.

The notion *construct* in language assessment typically refers to an abstract or theoretical concept or trait that exists independently and is hypothesised to underlie learning behaviours or language productions, for example (Bachman & Palmer, 2010). Theories (where available) can be useful for identifying the nature of the construct to be measured. The types of language and social constructs that language teachers tend to be interested in are intangible and not directly observable, and hence, cannot be directly measured. However, they can be inferred

from systematic and careful observation (e.g., the use of multiple well-crafted assessment items and tasks to elicit performance).

Psychometrics has influenced large-scale standardised tests used for estimating language proficiency and for selecting and admitting people to jobs or programmes of study, for instance. In educational assessment, national or state-standardised educational tests are typically administered to students to assess their literacy and numeracy skills objectively. In some settings, the results are often compared across schools, school districts, and states.

Psychometric validity

The discussion above has touched on some concepts of validity as they pertain to assessment. The notion of validity relates to the meaningfulness of the insights that are obtained about an individual's language ability based on their performance, and, in turn, the resulting inferences that are made about what they know, can do, or their skill level (see Chapelle, 2012, 2021; Kunnan, 2018; Lissitz, 2009). Over the past several decades, the fields of psychometrics, educational measurement, and assessment have focused on defining and exploring validity and validation (i.e., generating evidence to demonstrate validity) through the use of various approaches and frameworks. There is, as yet, no consensus on what validity means and many different interpretations of the term exist (Phakiti & Isaacs, 2021). For example, as outlined below, there is a longstanding debate about whether validity is a property of the test and whether the consequences of test score use and interpretation should be considered separate from or integral to the concept of validity (see Newton & Shah, 2016). Through the contributions of several scholars, including Cronbach and Meehl (1955), Messick (1989), and Kane (1992), to name just a few, the definition of validity and guidelines for validation have evolved, including in different editions of the *Standards* (AERA, APA, & NCME, 2014).

Validity was formally defined in the psychological literature, which began describing assessments in the 1930s (Lissitz & Samuelsen, 2007). The classic definition of "the degree to which a test or examination measures what it purports to measure" (Ruch, 1924, p. 13) suggested that the original focus of validity theory was on the test itself. That is, validity was believed to be an inherent, static property of a test. This concept was widely adopted and unchallenged until the early 1950s. From the 1960s, the original view of psychometric validity gradually shifted by recognising that no human assessment method is perfect (e.g.,

observations of behaviours or performance are fallible). However, it can be refined through rigorous processes.

In the late twentieth century, Messick (1989) introduced the highly influential *unitary concept* of validity. He (1995) posited that "validity is not a property of the test or assessment as such, but rather of the meaning of the scores. These scores are a function not only of the items or stimulus conditions, but also of the persons responding as well as the context of the assessment" (p. 741). Built on Messick's, validity is now considered in relation to the interpretation and use of test scores rather than being a property of the test itself (e.g., Chapelle, 2021; Kane, 2013). Echoing this, the current version of the *Standards*, AERA, APA, and NCME (2014) defines validity as "the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (p. 11). The focus, thus, is not on any given test, but rather on how test scores are interpreted and how those inferences about ability are used to inform decision-making.

The 2014 version of the *Standards* also maintains that "statements about validity should refer to particular interpretations for specific uses" (p. 11), which implies that the context of test use should be considered in validation (i.e., the process of generating validity evidence for a given assessment). That is, when an established test that has undergone rigorous validation for a particular context and purpose is being proposed for use in a different context and/or for a different purpose than that for which it was initially intended, it needs to be validated for that new context and purpose (Kane, 2013).

Practical guide I

Consider validity as a multidimensional concept guided by the following points in sequence:

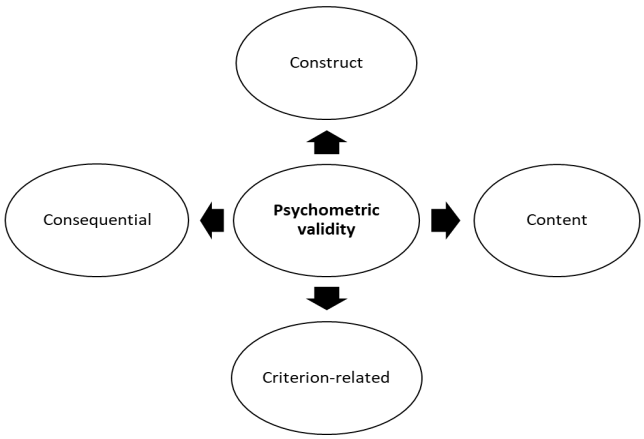
1. what is being assessed and how well can the assessment instrument do its job consistently and adequately for its intended purposes? To achieve this, teachers need to reflect on the steps they take in devising and using assessment methods (e.g., assessment instructions, conditions such as time allowance and test formats, and tasks or questions).
2. why is the assessment needed and what kind of inferences and decisions can be made on the basis of the assessment performance (e.g., test score)? For example, are students' scores rank ordered and compared? Are their scores compared against a set of language criteria or learning outcomes?

3.	what impact or consequences could the decision have on students (e.g., confidence, motivation, pass-fail, graduation, employment), on other educational stakeholders (e.g., parents, teachers), and on society at large (e.g., is the student suitable to perform a given job in terms of language proficiency and other factors?

Types of psychometric validity

Figure 2 summarises common types of validity in assessment that language teachers should apply to develop and improve effective classroom assessment practice, each concept of which will be elaborated on below. While face validity is not considered part of psychometric validity, it is included at the end of this section because it is an intuitive concept that reflects stakeholders' impressions of validity.

Figure 2. Types of psychometric validity



Construct validity

Construct validity can initially be understood in relation to the definition of constructs. That is, teachers first query the extent to which an assessment accurately and adequately measures the underlying construct they are targeting (e.g., language proficiency, reading ability, grammatical knowledge). They then move beyond that to reflect on whether the inferences made as a result of test scores or observations are reasonable assumptions to make based on students' performance manifestations on the (often unobservable) construct being targeted.

Messick (1989) views construct validity as the most important type of validity, as it is overarching in that it encompasses the other types of validity described below. This is a widely held view in educational and language assessment. Two notions that are related to construct validity are:

1. C*onstruct-irrelevance variance*, which denotes the influence of one or more variables that are extraneous to an assessment that can unduly influence the resulting performance and/or scoring outcome (Messick, 1995). For example, an online speeded essay writing test for L2 Chinese students might confound essay writing ability with other factors, such as the ability to shift from an English keyboard to select Chinese characters. Therefore, the assessment might not only reflect how well test-takers can write due to their computer literacy, as reflected in the assessment criteria, but also their ability to type rapidly—a potential source of construct-irrelevant variance.

2. *Construct underrepresentation*, defined as insufficient or incomplete inclusion of relevant content, items or tasks for eliciting information about a student's ability in light of the focal construct (Messick, 1995). In a classroom setting, performance samples (e.g., tasks or items) must adequately represent all aspects of the curriculum to ensure comprehensive assessment. For example, a pronunciation construct and curriculum target individual vowel and consonant sounds, word stress, rhythm, and intonation. If an assessment that seeks to make inferences about students' pronunciation ability only formally assesses individual sounds, it is underrepresenting the pronunciation construct.

Content validity

Content validity refers to the inclusion of relevant, representative, and sufficient samples of student behaviours or performances in an assessment in a way that comprehensively samples from a wider content domain or curriculum. Thus, it has an inverse relationship with construct underrepresentation in that the more content validity is weakened, the more the assessment fails to represent its intended content domain. The selection of content and tasks could be informed by theory (e.g., regarding the constituent components of the focal construct), hypotheses, and/or curricular objectives. An assessment that elicits a range of behaviours comprehensively representing the domain to which the performance seeks to be extrapolated will produce better content validity than one that elicits fewer and less representative behaviours (i.e., a construct underrepresentation

scenario). To maximise content validity, teachers need to select an appropriate and sufficient number of items from the content domain to which the assessment is supposed to extrapolate; however, this can be challenged by practical constraints, such as limited time and student or test-taker fatigue. There should be test specifications detailing the relevant domain that should be sampled from when developing the test. This can then serve as a benchmark for evaluating task and content coverage.

Criterion-related validity

Criterion-related validity refers to the relationship between test scores and some established external measure, such as scores from another test thought to measure a similar construct or real-world performance in a relevant domain that the test scores aim to predict. Two types of criterion-related validity are distinguished.

1. *Concurrent validity* refers to the extent to which test scores are related to an external criterion—a well-established measure that reflects the same or a closely related construct. This criterion might be another test believed to assess a similar construct, or a real-world performance outcome that the test is intended to predict. When two test scores from equivalent test forms (based on the same specifications) taken by the same test takers are correlated, a strong positive correlation coefficient is obtained, providing evidence of strong reliability. If the test forms are administered concurrently, such a correlation may also reflect strong concurrent validity.

2. *Predictive validity* denotes the extent to which a test score can predict future behaviours or performance (e.g., language proficiency scores may be used as predictors of students' academic attainment; e.g., Isaacs et al., 2023). In classroom assessment, predictive validity may be examined by correlating midterm scores with final assessment scores (e.g., to examine the extent to which students who perform well in the midterm also perform well overall on the final, and vice versa).

There are some points to note when the criterion measure in a predictive validity study is some form of classroom assessment. Bonner (2013) emphasises that assessments developed by classroom teachers often fail to meet the quality standards necessary for effective criterion measures—that is, measures that accurately predict the language proficiency test in question. Moreover, because the test is administered to only a small number of students, outliers can skew the results. Conversely, there have been suggestions that a broader range of criterion

measures could be investigated in tandem, including potentially classroom assessments or teacher or student perceptions (Pearson, 2020). In other words, teacher assessments could complement other indices, with potential advantages to having more than one criterion measure.

Consequential validity

Consequential validity refers to the extent to which test developers and test users consider both the intended and unintended consequences of test use (and potential misuse) on individuals, institutions, and the wider society. The term consequential validity focuses on investigating the consequences of test score use and interpretation on stakeholders, institutions, and society at large (see Messick, 1989). This concept highlights the importance of critically examining how assessments impact stakeholders within the broader educational landscape and, to the extent possible, as educators and policymakers, taking steps to mitigate potential negative consequences.

In the teaching and learning arena, consequential validity is often realised in terms of washback. Washback refers to the influence of assessment on teaching and learning behaviours and foci, often in a classroom context (see Alderson & Wall, 1993, for a classic reference). Washback is more pronounced in high-stakes assessments and is often investigated in the context of changes to a test resulting from a change or even a wholesale reform of a test or assessment system (Kim & Isaacs, 2018). For example, a language proficiency test that, after not assessing speaking, introduces a speaking component in a new version of the test, may prompt classroom teachers to focus on speaking during instructional time, whereas previously speaking may have been sidelined (Wall & Horák, 2008). It could also spur school principals, students, and parents to demand greater speaking instruction, and in turn, introduce market opportunities for individuals and companies to offer out-of-hours tutoring for supporting speaking development (e.g., cram schools) or promote language exchanges. Teaching to the test and narrowing the curriculum to focus on test content and format, or, conversely, engaging in more communicative behaviours to prepare for a new interactional task type, are examples of possible washback effects—positive and negative washback, respectively. Positive washback tends to occur when an assessment task is authentic (i.e., is similar to what test-takers need to be able to do in a real-world scenario) and can promote language skills that can be retained and applied outside of the assessment context. Conversely, a test that requires test-takers to listen and repeat pre-recorded utterances, for example. In contrast, if an

assessment asks students to complete a conversation through multiple-choice questions, they will likely not have the opportunity to speak or improve their speaking skills during test preparation or the test itself, and are unlikely to be able to transfer that knowledge to real-life communicative situations — an example of negative washback. Teachers need to foster positive washback to the best of their ability, within the constraints of the assessment system they are working in (much of which will be out of their control). They need to find a balance between assessment-driven instruction that meets system and stakeholder pressures and providing rich learning opportunities to genuinely enhance a student's language proficiency beyond the test itself. It is worth noting that when a high-stakes assessment is involved, some focus on test-wiseness strategies or techniques for "cracking" the test is likely to receive emphasis (Bachman & Palmer, 1996). This can be a source of construct-irrelevant variance if, for example, a student performs well or poorly not because of their reading comprehension ability, but simply because they are a skilled multiple-choice test-taker.

Face validity

Face validity is not a component of psychometric validity because it is not empirically measured and has been de-emphasised in literature on assessment and validity over the past several decades (Bachman & Palmer, 1996). It, therefore, is not included in Figure 2. However, it is a concept that some teachers are familiar with and, hence, warrants a brief discussion here. Face validity refers to the appearance of an assessment that looks as if it is assessing what it claims to assess (i.e., it looks right; it appears as if it is measuring what it says it does) from the vantage point of test-takers—the most common group referred to in relation to this concept. However, other stakeholders could also potentially be panelled about their perceptions of an assessment in this regard. For example, administering a grammar test to students and using the resulting scores to infer their writing ability may have low face validity if test-takers are not convinced of the relevance of the test as a measure of writing. Brown and Abeywickrama (2010) rightly pointed out that if students feel an assessment is not assessing what it is supposed to, "this might affect their performance" (p. 35).

HOW TO ENHANCE LANGUAGE ASSESSMENT VALIDITY

In this final section, readers are invited to consider three overarching principles that may enhance the validity of language assessment practices, potentially for both assessment for learning and assessment of learning.

1. Ensure the alignment among the intended learning outcomes, teaching and learning activities, and assessment tasks. When the intended learning outcomes are specified and are not articulated too broadly, it is more apparent what the focus of classroom instruction and assessment should be. Any items or tasks that are formally assessed should be familiar to students or test-takers. Ideally, they will have learnt to engage in such language use or learning activities in the classroom without drilling or teaching to the test. Aligning curricular goals with instructional delivery and assessments is not successful without considering students' or test-takers' characteristics to inform assessment design. Without good alignment, various types of validity (e.g., construct, content and criterion-related validity) cannot be sufficiently and appropriately addressed.

2. Recognise and follow key stages of assessment development and use (see e.g., Bachman & Damböck, 2017; Green, 2020). Teachers need to be able to form teams to work together on developing or refining an assessment due to the various stages, outlined as follows:

    a. The first stage is *planning*, including identifying the purpose of the assessment, focal construct, budget, human resources, and timeframe to complete. Planning allows a critical and realistic discussion of test development and of current limitations or challenges.

    b. The second stage is *developing*, including writing up specifications for the overall assessment, which may involve different task or item types, and providing a blueprint of those tasks or item types. The assessment may, in part, be based on a needs analysis or appraisal of students' language use demands in real-world settings or in reference to curricular goals.

    c. The third stage is *piloting*, including obtaining reviews of the assessment specifications from other stakeholders (e.g., teaching colleagues) and trialling items or tasks with other teachers or students that have similar characteristics to the target users.

    d. The fourth stage is *administering* or actually deploying the assessment that was specified in the second (development) stage, taking into account any modifications that were made as a result of the third (piloting) stage. This includes implementing administrative procedures that take into account logistical matters (e.g., implementing proctoring

to ensure smooth assessment delivery and mitigate cheating behaviours).

e. The final stage is *scoring and decision making*, including using objective or subjective scoring and performance reporting to students, ideally including meaningful, developmental, and individualised feedback provision.

3. Actively continue professional development in relation to language assessment in the interest of improving assessment literacy or an understanding of the knowledge and principles that teachers need to perform assessment activities to the best of their ability (Inbar-Lourie, 2017). Several language assessment books, such as Bachman and Damböck (2017), Brown and Abeywickrama (2010), Fulcher (2024), Ockey (2024), and Green (2020) as well as articles in academic journals such as *Language Testing, Assessing Writing* and *Language Assessment Quarterly,* present key considerations in language assessment design supported by research findings. These resources discuss ways to enhance the validity of assessments used in classroom settings, among other things. They also help teachers become more critical consumers of assessment and curricular materials, for example, in critically appraising a test provider's claims about what they are measuring, at what target proficiency level (and claims about CEFR alignment often should not be taken at face value), and how they go about doing so. We recognise that some of these resources require purchase or are behind a paywall requiring a subscription.

Even when accessed, journal articles may be written in a way that is conceptually inaccessible to teachers (Marsden & Kasprowicz, 2017). Open Accessible Summaries in Language Studies (OASIS), an open access digital resource (Marsden et al., 2018), are one-page accessibly written summaries of key research findings from journal articles in applied linguistics, including *Language Testing*, which encourages authors to craft a summary of their study once their article has been accepted for publication for uptake by a wider audience, including language teachers. Although relatively few assessment-specific articles are available, compared to other areas in applied linguistics (e.g., second language acquisition, psycholinguistics), those that do exist are an excellent, time-efficient way for teachers to engage with research. Summaries may be provided in multiple languages, including sign language, and some journals encourage or even require author engagement with them.

There is growing evidence that such summaries have the potential to widen outreach and engagement while also improving citation metrics (McKinley et al., 2025).

In some cases, OASIS summaries also link to open-access research instruments or tools that may also be teacher-relevant. For example, Isaacs et al.'s (2018) OASIS summary charts the development of a formative assessment tool for English for Academic Purposes teachers, focusing on international university students' English comprehensibility. A hyperlink to the tool with guidance for teachers on how to use it is available open access through the companion resource, the IRIS Repository of Instruments for Research into Second Languages (Marsden et al., 2016). OASIS is also collaborating with the new TESOLgraphics project, which summarises secondary research for teachers in a digestible format using infographics. Spoken abstracts and blogs are also becoming increasingly common to draw on a broader audience and make research more accessible. It is likely that, with the growing emphasis on open science in applied linguistics (Liu, 2023) and language testing (Kremmel & Isbell, 2024), there will continue to be an uptick in such sharing practices, particularly when funders or journals require or incentivise them.

Assessment methods, scientific knowledge, and technologies evolve, making it important to engage in continual and active updates to keep current. Consulting credible resources that promote ethical and fair language assessment practice is important. A few relevant resources include but are not limited to *Tools to Enhance Assessment Literacy* (*TEAL*) *for teachers of English as an additional language* (Tools to Enhance Assessment Literacy, 2025), the *ILTA Code of Ethics* (International Language Testing Association, 2018), which has been translated into several world languages, and ILTA guidelines for practice (International Language Testing Association, 2024), and Code of fair testing practices in education by the Joint Committee on Testing Practices (2004).

Another way to update knowledge in language assessment is to take training through short courses or pre- or in-service teacher training programmes, participate in public seminar series on assessment often run through university research centres, which may be offered on-line for wider audience capture, or take part in a Massive Open Online Course (MOOC) by a credible and ideally well-established provider targeting teacher professional

development and assessment literacy (e.g., British Council). The Teachers' Assessment Literacy Enhancement (TALE) project is an example of an open access language assessment professional development resource for teachers (Tsagari et al., 2018). New test reviews published in the journal Language Testing are now available with open access, allowing stakeholders and the general public to view and download the critiques without restriction. The growing emphasis on open science in applied linguistics and higher education (Marsden, 2019) suggests that other tools and innovations are likely to emerge in the years to come.

4. Partially overlapping with the third point above, join the global online language assessment community. LTEST-L is an unrestricted listserv for language assessment researchers and practitioners, including teachers from around the world with an interest in language assessment. Postings include substantive discussions of current issues in language assessment, which teachers sometimes initiate regarding practical matters, in addition to language assessment job advertisements, workshops, training courses, conferences, and announcements of field-relevant funding opportunities or prizes. Membership in professional language assessment associations, many of which host conferences, is another way to stay informed about activities and engage in networking. Although some organisations, such as ILTA, charge membership fees, several ILTA-affiliated regional associations currently offer free membership (e.g., ALTAANZ in Australia/New Zealand, EALTA in Europe, AALA in Asia, etc.) and are eager for member engagement (Taylor, 2026).

CONCLUDING REMARKS

This article aims to provide an accessible, theoretically informed discussion of psychometric validity, its implications for teachers, and strategies for further developing understanding and shaping practice. Assessment is not only at the centre stage of classroom teaching and learning but is part of a broader educational ecosystem. It ideally needs to be tailored for particular assessment needs and contexts and evaluated for those specific uses and contexts (Kane, 2013). We hope that language teachers reading this article will now have a practical understanding of what psychometric validity is and can see how the principles could apply to teacher-developed assessments and to critically evaluating external, often standardised language assessments.

Practical guide II

1. Develop a blueprint or plan that specifies the language skills or learning outcomes to be assessed and how they are to be assessed. This is similar to designing a house before actually constructing it, using a robust, well-thought-out plan with sufficient detail of what the house will look like once it is built.
2. Produce items or tasks to elicit enough evidence of meaningful performance or ability from the curricular targets or intended learning outcomes being sampled to avoid errors of reliance on a single item type that could increase the probability of chance findings. Where possible, elicit multiple snapshots of performance using a variety of methods, which should provide a more solid basis for making deductions about a learner's language ability.
3. Develop clear and transparent instructions for students to follow that are ideally as succinct as possible while being comprehensive enough to inform learners about what they need to do unambiguously. Details such as mark allocation, planning time/time on task, among others, should be provided. Clear administrative procedures and instructions can mitigate misunderstandings and ensure fairer practices for all.
4. Consider the consequences that interpreting the test scores and making decisions on that basis will have on students and other stakeholders. If the decision is high-stakes or impactful, more effort, time, and ideally resources and rigour will be needed for test development, preparation, and decision-making processes.

ACKNOWLEDGEMENTS

I would like to thank Aek Phakiti, Editor of the *University of Sydney Journal of TESOL*, for his substantial contribution to the conceptualisation and early drafting of this work. I was inspired by his drive to demystify opaque assessment concepts and make them more accessible and relevant for classroom teachers, as reflected in the goal of this article. In accordance with journal policy, Aek could not be included as an author, but I gratefully acknowledge his intellectual input.

AUTHOR

Talia Isaacs is Professor of Applied Linguistics and TESOL and Head of Research in her department at the UCL Institute of Education, University College London and co-editor of the international refereed journal *Language Testing*. Her research interests in language assessment are eclectic, including both standardised assessments and developing pedagogically-oriented tools to help learners calibrate views about their own performances through feedback provision. She has designed or led modules in applied linguistics and language assessment at universities in the UK, the US, and Canada and regularly disseminates her work to international audiences. Her collaborative research on assessment has been featured in various media sources, including Times Higher Education, The New York Times, the Canadian Broadcasting Corporation, and BBC Local Radio.

REFERENCES

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME) (2014). *Standards for educational and psychological testing*. American Psychological Association.

Joint Committee on Testing Practices (2004). *Code of Fair Testing Practices in Education*. https://www.apa.org/science/programs/testing/fair-testing.pdf

Bachman, L. F. (2004). *Statistical analyses for language assessment*. Cambridge University Press. http://doi.org/10.1017/CBO9780511667350

Bachman, L. F., & Damböck, B. (2017). *Language assessment for classroom teachers*. Oxford University Press.

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford University Press.

Baird, J-A, Johnson, S., Hopfenbeck, T. N., Isaacs, T., Sprague, T., Stobart, G., & Yu, G. (2016). On the supranational spell of PISA in policy, *Educational Research*, 58(2), 121–138, http://doi.org/10.1080/00131881.2016.1165410

Bonner, S. M. (2013). Validity in classroom assessment: Purposes, properties, and principles. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 87–106). Sage. http://doi.org/10.4135/9781452218649.n6

Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511490026

Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices*. Pearson Longman.

Brown, J. D., & Trace, J. (2017). Fifteen ways to improve classroom assessment. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning Volume III* (pp. 490–505). Routledge.

Burgess, S. (2015, March 9). *Gender, teacher assessment and stereotypes*. CMPO Viewpoint. Centre for Market and Public Organisation. https://cmpo.wordpress.com/2015/03/09/gender-teacher-assessment-and-stereotypes/

Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (pp. 257–270). Springer. https://doi.org/10.1007/978-3-319-02261-1_19

Chapelle, C. A. (2012). Validity argument for language assessment: The framework is simple… *Language Testing*, *29*(1), 19–27. https://doi.org/10.1177/0265532211417211

Chapelle, C. A. (2021). *Argument-based validation in testing and assessment*. Sage.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302. https://doi.org/10.1037/h0040957

Cui, Y., Meng, Y., & Tang, L. (2025). Reconsidering teacher assessment literacy in GenAI-enhanced environments: A scoping review. *Teaching and Teacher Education*, *165*, 105163. Advance online publication. https://doi.org/https://doi.org/10.1016/j.tate.2025.105163

Fulcher, G. (2024). *Practical language testing* (2nd ed.). Routledge.

Guardian (2021). *Let's not return to flawed exams. We have better ways to assess our children*. https://www.theguardian.com/commentisfree/2021/aug/15/lets-not-return-to-flawed-exams-we-have-better-ways-to-assess-our-children

Gonzalez, R. A., & Firestone, W. A. (2013). Educational tug-of-war: Internal and external accountability of principals in varied contexts. *Journal of Educational Administration*, *51*(3), 383–406. https://doi.org/10.1108/09578231311311528

Green, A. (2018). Assessment of learning and assessment for learning. In J. I. Liontas (Ed.), *TESOL encyclopedia of English language teaching*. John Wiley and Sons.

Green, A. (2020). *Exploring language assessment and testing: Language in action* (2nd ed.). Routledge. https://doi-org.ezproxy.library.sydney.edu.au/10.4324/9781003105794

Harding, L. (2026). Utopian and dystopian visions: Steering a course for the responsible use of artificial intelligence in language testing and assessment. *Language Testing*, *42*(4), 561–575. https://doi.org/10.1177/02655322251350717

Harding, L., & Kremmel, B. (2017). Teacher assessment literacy and professional development. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 413–427). De Gruyter Mouton.

Hu, R., & Trenkic, D. (2021). The effects of coaching and repeated test-taking on Chinese candidates' IELTS scores, their English proficiency, and subsequent academic achievement. *International Journal of Bilingual Education and Bilingualism*, *24*(10), 1486–1501. https://doi.org/10.1080/13670050.2019.1691498

Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, I. G. Or & S. May (Eds.), *Language testing and assessment* (3rd ed.), pp. 257–270. Springer.

International Language Testing Association (ILTA). (2018). *ILTA code of ethics*. https://www.iltaonline.com/page/CodeofEthics.

International Language Testing Association. (2024). *ILTA guidelines for practice in English*. https://www.iltaonline.com/page/ILTAGuidelinesforPractice

Isaacs, T. (2016). *Assessing speaking*. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 131–146). De Gruyter Mouton.

Isaacs, T., Hu, R., Trenkic, D., & Varga, J. (2023). Examining the predictive validity of the Duolingo English Test: Evidence from a major UK university. *Language Testing*, *40*(3), 748–770. https://doi.org/10.1177/02655322231158550

Isaacs, T., Trofimovich, P., & Foote, J. A. (2018). Developing a user-oriented second language comprehensibility scale for English-medium universities. *Language Testing*, *35*(2), 193–216. https://doi.org/10.1177/0265532217703433

Jones, N., & Saville, N. (2016). *Learning-oriented assessment: A systemic approach*. Cambridge University Press.

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*, 527–535. https://doi.org/10.1037/0033-2909.112.3.527

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*, 1–73. https://doi.org/10.1111/jedm.12000

Kim, H., & Isaacs, T. (2018). Teachers' voices in the decision to discontinue a public examination reform: Washback effects and implications for utilizing tests as levers for change. In D. Xerri & P. Vella Briffa (Eds.), *Teacher involvement in high stakes language testing* (pp. 263–282). Springer.

Katz, A. (2012). Linking assessment with instructional aims and learning. In C. Coombe, P. Davidson, B. O'Sullivan & S. Stoynoff (Eds.), *Cambridge guide to second language assessment* (pp. 66–73). Cambridge University Press.

Kremmel, B., & Isbell, D. R. (2024). Open Science practices in language assessment: Introducing the special issue. *Language Testing*, *41*(4), 697-702. https://doi.org/10.1177/02655322241264092

Kunnan, A. (2018). *Evaluating language assessments*. Routledge.

Leung, C. & Rea-Dickins, P. (2007). Teacher assessment as policy instrument: Contradictions and capacities. *Language Assessment Quarterly*, 4(1), 6–36. http://doi.org/10.1080/15434300701348318

Lissitz, R. W. (Ed.) (2009). *The concept of validity: Revisions, new directions, and applications*. Information Age Publishing.

Lissitz, R. W., & Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher*, *36*(8), 437–448. https://doi.org/10.3102/0013189X07311286

Liu, M. (2023). Whose open science are we talking about? From open science in psychology to open science in applied linguistics. *Language Teaching*, *56*(4), 443–450. https://doi.org/10.1017/S0261444823000307

Lopez, L. M. & Pasquini, R. (2017). Professional controversies between teachers about their summative assessment practices: a tool for building assessment capacity. *Assessment in Education: Principles, Policy & Practice*, *24*(2), 228–249. http://doi.org/10.1080/0969594.2017.1293001

McNamara, T. F. (1996). *Measuring second language performance*. Longman.

Marsden, E. (2019). Open science and transparency in applied linguistics research. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. https://doi.org/https://doi.org/10.1002/9781405198431.wbeal1493

Marsden, E., Alferink, I., Andringa, S., Bolibaugh, C., Collins, L., Jackson, C., Kasprowicz, R., O'Reilly, D., Plonsky, L. (2018). *Open Accessible Summaries in Language Studies (OASIS)* [Database]. Available: https://www.oasis-database.org

Marsden, E., & Kasprowicz, R. (2017). Foreign language educators' exposure to research: Reported experiences, exposure via citations, and a proposal for action. *The Modern Language Journal*, *101*(4), 613–642. https://doi.org/https://doi.org/10.1111/modl.12426

Marsden, E., Mackey, A., & Plonsky, L. (2016). The IRIS Repository: Advancing research practice and methodology. In A. Mackey & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS Repository of Instruments for Research into Second Languages* (pp. 1–21). Routledge. https://doi.org/10.4324/9780203489666

McKinley, Jim, Baranowski, Mariusz and Cichocki, Piotr. (2025). Do open access plain-language summaries increase engagement with research? *Applied Linguistics Review*. Advance online publication. https://doi.org/10.1515/applirev-2024-0269

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). American Council on Education/Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

Newton, P., & Shaw, S. D. (2014). *Validity in educational and psychological assessment*. Sage. https://doi.org/10.4135/9781446288856

Ockey, G. J. (2024). *Introducing second language assessment*. Cambridge University Press.

OECD. (2021). *Education policy outlook 2021: Shaping responsive and resilient education in a changing world*. OECD Publishing. https://www.oecd.org/en/publications/education-policy-outlook-2021_75e40a16-en.html

Pearson, W. S. (2020). The predictive validity of the Academic IELTS test: A methodological synthesis. *ITL - International Journal of Applied Linguistics*. https://doi.org/https://doi.org/10.1075/itl.19021.pea

Phakiti, A., Isaacs, T. (2021). Classroom assessment and validity: Psychometric and edumetric approaches. *European Journal of Applied Linguistics and TEFL*, *10*(1), 3–24.

Pitts, R. T., & Naumenko, O. (2016). The 2014 Standards for educational and psychological testing: What teachers initially need to know. *Working Papers in Education*, *2*(1), 1–6.

Poehner, M. E., & Infante, P. (2017). Dynamic assessment in the language classroom. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 275–290). De Gruyter Mouton.

Popham, W. J. (2017). *Classroom assessment: What teachers need to know* (8th ed). Pearson.

Ruch, G. M. (1924). *The improvement of the written examination*. Scott, Foresman.

Sato, M., Chong, S. W., Aktar, T., Cowell, J., Kong, M. S., & Shaahdadi, M. (2024). Creating and sustaining a platform for researchers and teachers to communicate: An example of TESOLgraphics. *Innovation in Language Learning and Teaching*. Advance online publication.

Shohamy, E. (2017). Critical language testing. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (pp. 441–454). Springer. https://doi.org/10.1007/978-3-319-02261-1_26

Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, *29*(1), 21–36. https://doi.org/doi:10.1017/S0267190509090035.

Taylor, L. (2026). What does it mean to be a member of a professional association in our field? *Language Testing*, *43*(1).

Tools to Enhance Assessment Literacy. (2025). *Tools to enhance assessment literacy (TEAL) for teachers of English as an additional language* https://teal.global2.vic.edu.au/

Tsagari, D., Vogt, K., Froelich, V., Csépes, I., Fekete, A., Green A., Hamp-Lyons, L., Sifakis, N., & Kordia, S. (2018). *Handbook of assessment for language teachers*. Retrieved from: http://taleproject.eu/

Turner, C. E., & Purpura, J. E. (2016). Learning-oriented assessment in second and foreign language classrooms. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 255–273). De Gruyter Mouton.

Wall, D., & Horák, T. (2008). The impact of changes in the TOEFL examination on teaching and learning in Central and Eastern Europe: Phase 2, coping with change (ETS Research Report No. RR-08-37). ETS. https://www.ets.org/research/policy_research_reports/publications/report/2008/hspw.htmlAuthor.html

Making the Connections: A Teacher Educator's Perspective on Compatibility of English Language Teaching Methods

Mark Fraser[1]

University of Wollongong

ABSTRACT

This article is based on my keynote presentation at the *16th University of Sydney TESOL Research Colloquium* in 2024. Teaching about the range of methods that characterise our field is standard in any Teaching English to Speakers of Other Languages (TESOL) education course. Common texts provide teacher educators with guidance on how to raise learners' awareness of the origins of methods, their similarities and differences, and the advantages and challenges of using these methods to teach English to linguistically diverse learners. However, while such texts and other resources help develop an understanding of various TESOL methods, I found that very few provide guidance on how to teach our learners about the compatibility of these methods with their own contexts. Just as there is no correlation between proficiency levels in English language macro skills, experience in teaching English as an additional or foreign language does not guarantee sufficient levels of proficiency in English language teaching methods. In this session, I will share with you my self-study story on ways I approach and develop the teaching of TESOL methods to a linguistically diverse cohort of postgraduate students. At the core of my approach is the adaptation of a critical

[1] Correspondence: Mark Fraser: mfraser@uow.edu.au; https://orcid.org/0000-0003-3331-560X

ISSN 1834-4712 (Online).

DOI: 10.5281/zenodo.17590992

Citation: Fraser, M. (2025). Making the connections: A teacher educator's perspective on compatibility of English language teaching methods. *University of Sydney Journal of TESOL*, *4*, 163–178. https://doi.org/10.5281/zenodo.17590992

reflection framework to inform course design, which enables students to make stronger connections between theory and practice. My journey as a teacher educator continues, and I will discuss the implications of this approach for helping our students make connections by exploring the compatibility of the content, making it both meaningful and transformational for their own teaching contexts.

Keywords: contextualisation, critical reflection, professional identity, reflective practice, teacher education, TESOL methods

INTRODUCTION

Teacher educators often inhabit liminal spaces, negotiating between research, practice, and teaching-intensive responsibilities. While much of the scholarship foregrounds the contributions of research-focused academics, there is value in considering how teaching-focused academics engage with, and make connections to, research in ways that shape student learning and professional growth. The aim of the keynote address I gave at the 2024 University of Sydney TESOL Research Colloquium was to share my own journey as a teacher and teacher educator, and to describe a reflective framework I use to help postgraduate TESOL students work out what will and will not fit their own teaching situations: acknowledging the things that inform our teaching practice and teacher identity. While it was a big honour to be invited to speak at the event, I felt intimidated; the impostor syndrome indicator was ticking up. However, putting those feelings aside, I reflected deeply on who I am now and what I have done to get to this point. I am one of those academics defined as teaching-intensive, meaning that around 80% of my academic work involves teaching. The requirement to undertake significant research is less for me than for many of my peers. Can you see why I sometimes feel like an impostor among researchers and budding researchers? As mentioned, interpretations of perspectives and our interactions with these interpretations are core to understanding our value and the value of our contributions to our social world.

Thus, it occurred to me that while we often hear from academics with well-established research identities, it is also important to hear from those whose job it is to make connections between the work of researchers and those who are impacted or those who will use the research to impact others—the teacher educator. As a teacher educator, I share my story of how I help my students make meaningful connections in my TESOL methods subjects and throughout my teaching. Core to my story are two main factors: the role the process of negotiating

our professional identities plays in teacher education, and the value of creating a reflective space to explore this process as we develop our knowledge and understanding of TESOL methods. My story does not rely heavily on a specific research methodology, but it is informed by research. You could say that it reflects the spirit of ethnography, the study of cultural patterns and behaviours to understand the complex social world in which I live. Thus, here is my story, comprised of 4 episodes.

EPISODE 1: THE TEACHER: YOU'RE DR WHO?

We all have a starting point in our teacher education journey. Here is mine. The starting point for my story is my broad research interest in exploring professional identity. It has come to underpin all that I do in my teaching. This was the research topic of my doctoral studies. The inclination to explore this area has always fascinated me, possibly because I am drawn to understanding what drives people's actions and what shapes them as individuals. Since deciding to follow a career in teaching, the number one participant in all my observations has been me. Thus, in the spirit of ethnography and following the theme of professional identity, let me bring you up to speed on who I am.

I had never considered teaching as a career, let alone one in academia. My path beyond high school was limited due to my failure in the Higher School Certificate (HSC), the final exams we take at the end of our high school days in NSW. English was my weakest subject. This result defined me and shaped my perception of my intelligence for many years. It seemed I was destined for something less than greatness, but I was lucky to have whatever came my way, so I stuck with it.

So, I spent 8 of my post-HSC years working for a bank, until my sister secretly sent off a university application on my behalf. Unknowingly, stars began to align. Three years later, I earned a business degree with a major in tourism. My dream was to travel the world for the rest of my life and what better way than to have a job in tourism. Moreover, when I graduated, I had a plum government job waiting for me with Tourism NSW, but a little advertisement caught my attention to teach English in Taiwan. What better way to improve my English than to teach it!

After a coin toss, and in the blink of an eye, I was on my way to Taiwan. Virtually untrained, definitely inexperienced, and still with the stigma of having failed English in my final year at school (though I did leave university with a distinction average), I was standing at the front of a classroom, filled with young learners and

their parents, ready to be taught English by this newly arrived native speaker teacher. My one semester of postgraduate study in TESOL, which I managed to fit in just before leaving Australia, did not cover this scenario! How could that Derewianka SFLish text and context class help with teaching present simple tense dialogue about things we eat for breakfast? I was not making any connections. Halfway through the first class, I politely excused myself and found the first cubicle in the men's toilet where I collapsed on the floor in absolute fear. What the hell had I got myself into? What is going to get me out of this one?

Let us stop here to make some connections.

I present the pieces to my identity puzzle: Me with my self-assigned label of school failure. Worked in a bank. Completed an undergraduate degree, beyond expectations. Added a postgraduate qualification in TESOL to my resume. Which of these had the most significant impact on me, do you think, while I was on the floor of the cubicle, reflecting on my identity as an English teacher in a Taiwanese conversation school for mostly unmotivated adolescents? Over the course of eight years, I worked in an environment where quality service was at the core of every activity we did. Checks and balances ensured that customers' needs were always met effectively, efficiently, and to a high standard. If this were absent in my DNA after my time in the bank, I would have been on the next plane out of Taiwan. Despite the academic success and the brief training, it was this experience that pulled me through. That and my dad's parting words, someone who, as a metal fabrication teacher with a healthy level of scepticism of universities, told me that to be a good teacher, "you gotta know what you are talking about".

So, I made the commitment to do my best, which meant learning how to teach English. In this context, English language teaching at a cram school in Taiwan was heavily influenced by the traditional approach, and many local Taiwanese teachers adhered closely to the teacher-centred approach, undertaking the bulk of the teaching work with a heavy reliance on Mandarin as the language of instruction. As a native English speaker, I was the novelty act, coming in and essentially being the clown who built on the work of the Taiwanese teacher. What made this school unique, too, was that parents could sit in the back and learn English alongside their children. Most of the time, the kids were asleep, their heads on their desks, while the parents diligently took their notes. It was a case of 2 for the price of 1 as the parents were more motivated than the students to learn English. Grammar structures were taught through choral drilling with accuracy one week, only to be forgotten the following week. It was a time of very steep learning on how to teach

English. I developed my own values regarding methods of teaching and learning English as a foreign language, tailored to this specific context, believing that students needed to have a choice. An understanding of the motivational factors that contribute to their learning needs to be identified and acknowledged. Furthermore, they needed to work at their own level. I began to reflect on the communicative approach to teaching language and a more autonomous approach to learning language. However, I still lacked the depth of knowledge about appropriate pedagogy at that stage. I had huge arguments with the local teachers who told me that the students should not be taught my way, that the Taiwanese knew the right way to teach. Maybe they were right. I was still green.

Let me fast-forward now, past the two years I spent working in an English-medium instruction kindergarten, still in Taiwan, where I oversaw 16 three-year-olds who spoke no English. And with no qualifications in early childhood education. Back in Australia, I pursued a Master of Education in TESOL, as I believed that to stick to this career path, I needed to gain more knowledge and qualifications. Let us skip over most of my eight years in Japan, where I worked at two universities, teaching English to mostly unmotivated students. Nevertheless, let us stop briefly towards the end of my time there. This is where I began my doctoral studies, marking the start of my journey to understand how our professional identities are shaped by a complex process of interpreting interactions within past, present, and future experiences. However, once again, my time in the bank informed the genesis of my doctoral studies. How can we provide learners with value for money? What makes a good teacher? Not just any teacher of English, but one who has committed to a career of teaching English in tertiary contexts, in my case, Japan. What informs the negotiation of professional identities of teachers of English in Japan?

Undertaking my doctorate at UOW, I was surrounded by significant theoretical forces, including Bakhtin, Bourdieu, Foucault, and ubiquitous sociocultural theory. All of which I struggled to grasp, though I did enjoy Foucault's writings on incarceration. Instead, I was drawn to symbolic interactionism. It is a theoretical framework informed by the work of Mead (1934), who was aligned with the University of Chicago. You will not find a great deal of his publications because he did not publish a great deal. Instead, he preferred to share his research through presentations and lectures. His theories fall within the field of social psychology, although there is little evidence that he defined himself as a social psychologist.

The central premise of symbolic interactionism is that social reality is constructed through the interactions and meanings that individuals assign to symbols,

language, and gestures in their everyday lives. It emphasises the importance of understanding how people interpret and give meaning to their experiences and how these interpretations shape social behaviour and relationships. Essentially, it suggests that human behaviour is influenced by the social context and the meanings individuals derive from their interactions with others. If you think you have heard something like this before, it is because a lot of what Mead shared was very similar to that of Vygotsky (Lantolf, 2000). There is no evidence that the two ever met, but comparing the principles of symbolic interactionism with those of sociocultural theory reveals a significant overlap. We could be mistaken in thinking that Mead and Vygotsky were playing a game of 'anything you can do, I can do better.' I will leave that argument for the theoretical purists on either side.

I was drawn to symbolic interactionism because it made sense to me and seemed appropriate for the world I wanted to explore. It seemed a nice fit for a sticky beak like me to explore the professional identities of teachers established in their careers of teaching English in Japan. Symbolic interactionism provided me with insight into how others perceive the teaching of English in specific contexts, specifically in the foreign language teaching context at a Japanese university. Through my research, I found that they drew on not only their own language learning experiences but also a complex matrix of other factors that informed their teaching practice. As part of this negotiating process, many were challenged by local customs and local approaches to teaching English and the education system. Context mattered. However, more than context, such as my own extensive banking experience, many non-educational factors also played a significant role in how they negotiated their identities as English teachers. From this, a spark was born.

Our professional identity influences the way we interpret our interactions with TESOL concepts as we grow and develop in our roles as English teachers. However, something was missing. How can we characterise this process of interpreting our interactions with the range of objects —social, physical, and abstract — that contribute to our growth? How were they making these meaningful connections?

EPISODE 2: THE CONTEXT: FINDING MY FOOTING

It is now time to put me into context. Let me set the scene. Allow me to transport you to the wonderful world of the University of Wollongong's TESOL postgraduate programme. Our student cohort is wide-ranging. Domestic and international. From the inexperienced to the very experienced. As is often the case in higher education, we move and teach between subjects within our TESOL program. One sunny day,

not too long ago, I was invited to take over the subject of Methods in Second Language Learning. Your standard methods subject is popular with many. A core component of several of our TESOL courses. A great opportunity, but one that I was apprehensive about accepting. We know that just because a language learner can speak well, their ability to write in the target language does not always correlate with their speaking ability. Similarly, just because I can teach English, and English is my first language, does not mean I know how to teach the methods that inform my teaching. Especially when it is not my area of research. I was filled again with doubt. We teachers often doubt our abilities, don't we? I was mentally dialling the identity crisis hotline. Not only about ways to teach a methods class in TESOL, but what did I know about TESOL methods that I could teach? More broadly, what was my area of expertise, what was my thing that I could bring to this subject? I work with colleagues who are experts in pronunciation, functional grammar, and neuroscience. What was my thing? I have a PhD in professional identity. Where is the value in that? It is not as sexy as, say, doing the dialogic approach the Wotring way (Wotring et al., 2024). Thus, when I was asked to teach methods, I had no idea what my area of expertise was at that stage that would inform my take on teaching methods and add value. The banking DNA was driving a need to maintain a sense of quality and service. Challenge accepted! I always take on tasks that are beyond my capabilities. That is how we grow, I suppose.

So, my first semester of teaching methods was literally by the book. The Jeremy Harmer textbook 'The Practice of English Language Teaching, 5th edition' (Harmer, 2015). Well known through its many iterations. As is the case when subjects are passed on, I inherited teaching materials from the previous coordinator.

For the first semester, I was essentially parroting what was given to me without truly connecting with much of the content. It began by defining and explaining popular TESOL methods, such as communicative language teaching (CLT), task-based language teaching (TBLT), and content language integrated learning (CLIL) (Harmer, 2015). I had been aware of these concepts, which I incorporated into my teaching at various points in my career, dating back to my time in Taiwan. However, I still thought the depth of my conceptual knowledge was questionable. The textbook and other sources provided sufficient information to begin building conceptual knowledge and introducing students to the topics, but something was still missing – something was not just connecting. As teachers, we need to bring that content to life, help our students make connections, and, more importantly, ensure it is relevant to their contexts. I needed to examine the connections more closely to make them meaningful.

I attempted to create interactive lessons, informed by the methods we discussed, to demonstrate and evaluate, but something was still missing. What did they need to trigger that meaningful process? What a goofball! I thought. I was so focused on conveying the content that I overlooked the simple fact that I was teaching teachers, or people who aspired to be teachers. So stupid me, of course. TESOL teacher education courses like ours are great, focusing on what and how to teach English, but perhaps they are a bit light on the 'who'. Found my thing! Engaging their professional identity involves negotiating the objects that shape their self-perception as English teachers, how others perceive them, and how these interactions inform their teaching practices. Thus, that was the trigger that I needed. To then find that connection that brings about this process, and of course, I had it right in front of me: Critical reflection.

EPISODE 3: CAN I HAVE THE ROOM? MAKING THE CONNECTIONS IN THE REFLECTIVE SPACE

It was an area of research that I started as a casual academic to keep me from going balmy, taking on marking jobs for other academics. Fresh from submitting my thesis, I naively volunteered to mark essays in one of our preservice teacher education courses in our first-year subjects. Core to all the essay tasks was the requirement to reflect on various teaching concepts critically. After marking nearly 1000 essays over the course of one semester, I realised that the students could not write critical reflections. They wrote reflections, but they did not write critical reflections. Why couldn't they write a critical reflection? More puzzling, why not simply Google 'how to write a critical reflection'? This sparked my interest in examining critical reflection as a concept and exploring how learned knowledge can be demonstrated through a critical reflection lens, primarily through writing. They needed a useful tool or resource. Reflection was an assumed concept, something that everybody knew how to do. Reflection was a common element throughout many teacher education subjects, including TESOL. Despite this shared understanding and commonly used concept, there was very little guidance on teaching how to write a critical reflection or assess critical reflection writing. For UOW students, I needed to come up with something that I could teach them about both the concept and as a genre of writing. However, offer something that would not be clunky, that would not be a colour-by-numbers type of approach to writing. Something that would provide them with the factors to consider for inclusion, while also giving them the freedom to be creative. So, how do I differentiate between reflection and critical reflection? I am glad you asked.

Reflection is a complex process that involves interpreting events and experiences of the past in the context of the present, informing future action and behaviour (Blumer, 1986). According to Brookfield (2017), however, reflection is not, by definition, critical. Critical reflection, as a concept, is situated further along the continuum of reflection, with a greater emphasis on encouraging movement beyond recollections and observations towards analysis, action, and change. Described as a complex, rigorous, intellectual, and emotional enterprise (Rodgers, 2002, p. 844), it provides space to explore 'both personal and professional belief systems' (Larrivee, 2008, p. 343). While critical reflection's value and role in teacher education have been well documented (e.g., Körkkö et al., 2016; Liu, 2015; Ramlal & Augustin, 2020), less clear is how students can effectively demonstrate their ability to reflect critically on experiential learning in their writing.

The works of Brookfield (2017), Dewey (Dewey, 1910), and Schön (Schön, 2017) informed my understanding of critical reflection as a concept, in addition to the work of Ryan and Ryan (2014) and Bain et al. (2002) to help develop learning resources focused on critical reflection writing. While my conceptual understanding remains aligned with these reference points, I found that during this self-exploration, a fundamental tension emerged between my conceptual knowledge and its application in teaching and assessing critical reflection writing at the graduate level. Ryan's (2011) critical reflection writing framework, for example, draws on a systemic functional linguistic (SFL) perspective, comprised of four text types: recount, description, explanation, and discussion (Ryan, 2011, p. 104). Each text type has particular language features.

Teacher education programmes often lack a sufficient structure or framework to develop critical reflection that effectively contributes to transformative learning. Schön's (2017) popular concepts of reflection in action and reflection on action have been used widely in teacher education to raise awareness about the value of the relationship between critical reflection and teaching practice. Although the literature on critical reflection and suggested frameworks for developing reflective practices are growing, few provide clear guiding frameworks for developing writing practices that demonstrate critical reflection.

Viola! I came up with something very simple. Perhaps too simple, but this critical reflection framework (Fraser et al., 2024) enables me to clarify what I mean by critical reflection and what is expected when students write a critical reflection. It has three points, as all triangles do. Furthermore, each point represents key elements of a critical reflection. The process begins with an experience or episode,

something personal to them, relatable, and about which they have extensive knowledge. They are then asked to describe and explore that experience through the lens of theoretical concepts that relate to it. This provides greater depth to their experience, enabling them to gain a more nuanced understanding of that experience that aligns with the broader research literature. And then? Through that process of meaning-making, one starts to look to the future. How does this meaning-making contribute to the negotiation process of developing their professional identities? To develop their teaching practices that meet the expectations of others in their target teaching contexts? The critical reflection aims to prompt them to explore who they are in the present, drawing on something from the past, but to bring about a transformation that will propel them further into the future in a more developed way. Thus, that is my critical reflection.

That was the easy bit. The challenge has been to find common ground where we can share understanding and engage in conceptual discussions. It is taking me some time to find that ground, to build that ground – let me keep the metaphor going – to fertilise and nurture that ground. I am still getting there, but the tools, compost, and fertiliser that I have been using to work this ground consist of developing students' abilities to look for compatibility in what we are discussing. It has been a common remark that students who pass through our program, particularly the international students, enjoy learning and engaging with the new perspectives, and, only to return to their home contexts, feel they have little opportunity to implement this new learning (Pham & Saito, 2019). This issue is one that I have been trying to address in all my TESOL teaching. How to make what they do and learn here have meaning back there? How to make it compatible? A critical reflection thread throughout my subject design creates a productive reflective space that facilitates discussions that explore this issue.

Setting off down a rabbit hole, where I think I am still traversing, I explored that process, starting with a self-study of the critical reflection concept (Fraser, 2021). If I wanted my students to frame their discourse through critical reflection, I needed to gain a deeper insight into the concept that would inform my teaching methods. More importantly, the goal is for students to find value in the concept for their own personal growth and development.

Our interactions with the past help us to interpret the present, which, in turn, informs our future behaviour as teachers. My self-study provided a small but meaningful piece to my complex identity as a teacher educator by allowing me to pay closer attention to my teaching practices. Self-study in teacher education

practices (S-STEP) aims to explore and expand the knowledge base of teacher educators (LaBoskey, 2004). LaBoskey LaBoskey (2004, p. 839) argues that as educators of teachers, "we have a pedagogical responsibility to monitor our progress continuously; to check for discrepancies between our ideals and our practice [in ways that] … result in, or not, the reframed thinking and practice of our students and ourselves". While self-study is an established methodology for practitioner inquiry within language education scholarship, with the dual purpose of optimising practice and contributing to a grounded and public knowledge base of language teacher education, its presence at the graduate level is underdeveloped.

Now, you are probably wondering about the trustworthiness of such a study. Well, I am glad you asked. Being a self-- study, I was uniquely placed as both participant and researcher, which some may say challenges current notions of trustworthiness. However, having my back, Hamilton et al. (2020, p. 309) suggest that rather than focusing on the primary research relationship between the researcher and the academy, the focus should be on the relationship between the researcher's world in which they live and practice. I constantly triangulated my journal data and margin notes with the research findings, the analysis of reflective posts made by my students in a blog and in their assessment tasks, and the research literature around critical reflection writing. Further triangulation, common to S-STEP, such as working with a critical friend, would have been beneficial in strengthening the trustworthiness of my data interpretation and providing suitable models for collaborative self-study work. Yet, I am to have that conversation with my critical friend, so my work is still in progress.

From my dalliance with critical reflection as a concept, I found that conceptual knowledge requires significantly more time than a semester, which we already know. However, creating reflective space can lay some important foundations upon which knowledge can grow beyond the subject. This is something that is rarely covered in any textbook that informs teaching TESOL methods. The development of conceptual knowledge is a process that needs time. The contributing factors that determine the length of time are our abilities to make things compatible through contextualisation, developed through engaging in critical reflection.

Reflective space is core to fostering effective growth and meaningful transformation of a teacher's professional identity and teaching practice. The reflective space I identified in my subject design is characterised by the students'

abilities to interact with their interpretations of a range of social objects that would otherwise be constrained if they attempted to seek reflective practice during their regular teaching routines and environments. Removal from their typical contexts placed them in a kind of "zero room" environment, free from the weighty demands of teaching and administrative responsibilities usually constraining any form of deep, critical reflection on their teaching practices and who they are as teachers. Time constraints and pressures from work commitments limit a teacher's ability to reflect meaningfully in a way that contributes to change. Within this reflective space, students can explore choices based on theoretical concepts relevant to their own teaching environments. However, this space is not constraint- or pressure-free. Students often begin their learning experiences with some level of anxiety due mainly to doubts about their abilities to undertake work at a postgraduate level and about their language abilities to write to the same level of expectations. However, our students are also highly supported within this space to address these limitations. An important factor in my subject is the use of scaffolding to minimise the impact of these doubts and maximise the opportunity for participants to develop confidence as they engage with the theoretical concepts.

More significantly, this opportunity allows my students to integrate their new ideas and teaching practices, applying the newly discovered theoretical principles within their respective teaching environments. The Methods subject, with critical reflection at its core, allows students to explore opportunities to meaningfully and sustainably integrate their knowledge into their teaching practices. The aim is not to revolutionise the entire education system, but rather to find a small yet suitably compatible way to enhance their teaching practices, to provide more effective learning experiences for their students.

The design of my methods class provides a reflective space for students to explore concepts and theories relevant to their own teaching environments, without the constraints common to their teaching contexts. Instead of transition, the process of critical reflection aims to bring about regeneration in both professional identity and teaching practice. Reflective space draws together threads of their past to be used in negotiations in the present, shaping future behaviour and practices that meet the expectations of becoming an English teacher in their respective educational contexts.

EPISODE 4: AND SO: CHALLENGES, CONSIDERATIONS, AND CONNECTIONS, OH MY!

Thus, my story comes to a close. A story about how I try to awaken my students to looking more closely at how their stories are evolving, making connections to understanding in greater depth their own professional identities as teachers of English. It is one that I am sure is familiar to many. The story of a teacher becoming: the doubts, the challenges, the growth, the success, the discoveries (often serendipitously). For me, it was through reflection that I have gained both a deeper understanding and greater appreciation of who I am becoming as a teacher. The way in which my professional identity informs my practice helps my students make their own connections. Of course, we personalise our teaching, informed by the infinite process of negotiating our professional identities, the complex matrix where experiences, knowledge, and perceptions of past, present and future interact.

Teaching methods are more than amplifying the pages of a textbook. It requires exploration of compatibility to make meaningful connections with conceptual knowledge and the complexity of real-world classrooms. There are challenges ahead in teaching English that will impact TESOL methods in teacher education. The affordances and challenges of using AI as a helpful tool in the language classroom are readily apparent. Learning English with the temptation of these emerging and impactful tools helps make connections beyond the local context and community, which is a significant factor in our teaching identity. It also has the potential to change the character of our community. For the better? Who is to know? As I teach the universal aspects of methods, I remind my students also to consider the impact. Reflect on the connections that will be made as a result of their teaching.

CONCLUSION

The three fundamental principles that inform my TESOL methods subject design (and the design of all of my teacher education subjects) are as follows. First, contextualisation is important. It is not just to help explain the methods. Getting students to draw on their respective contexts to explore more deeply the affordances and challenges. This process of interpretation provides the meaning with which we engage with the world. Second, it makes the job of finding compatibility a little less challenging. The realisation is that it is not about changing the system and turning it upside down. It is about finding inroads. It is about

finding opportunities to make the smallest transformation or have the smallest impact, yet achieve a meaningful impact in whichever teaching and learning environment we are practising. Third, the space in which all this happens, the realisation of context and the identification of opportunities for compatibility, takes place in a reflective space. At the core of making all this connected is the process of critical reflection. A magical thinking tool of infinite potential. I have found that it helps my students make those connections more meaningful. The feedback I receive from students indicates that, from their perspective, it makes their learning experiences more meaningful. However, more importantly, they can then take with them the knowledge they gain here and apply it in a meaningful way in their own context. I think that is how I perceive my role as a teacher educator, as a teaching-intensive academic. To provide those connections between the theory and practice, connections characterised by compatibility, contextualisation, and reflective space. It must be meaningful for me first before I can make it meaningful for others.

AUTHOR

Mark Fraser teaches in both undergraduate and postgraduate courses, specialising in TESOL (Teaching English to Speakers of Other Languages) at The University of Wollongong. His main research interests include critical reflection, professional identity, and self-study in teacher education.

REFERENCES

Bain, J., Ballantyne, R., Lester, N., & Mills, C. (2002). *Reflecting on practice: Student teachers' perspectives*. Post Pressed.

Blumer, H. (1986). *Symbolic interactionism: Perspective and method*. University of California Press.

Brookfield, S. D. (2017). *Becoming a critically reflective teacher*. John Wiley & Sons.

Dewey, J. (1910). *How we think*. *Heath & Co.*

Fraser, M. (2021). Through you I see me: S-STEPping into a teacher educator's identity through critical reflection. In R. Yuan & I. Lee (Eds.), *Becoming and being a TESOL teacher educator* (pp. 207–226). Routledge.

Fraser, M., Wotring, A., Green, C. A., & Eady, M. J. (2024). Designing a framework to improve critical reflection writing in teacher education using action research. *Educational Action Research*, *32*(1), 43–59. https://doi.org/10.1080/09650792.2022.2038226

Hamilton, M. L., Hutchinson, D., & Pinnegar, S. (2020). Quality, trustworthiness, and S-STTEP research. In J. Kitchen, A. Berry, S. M. Bullock, A. R. Crowe, M. Taylor, H. Guðjónsdóttir, & L. Thomas (Eds.), *International handbook of self-study of teaching and teacher education practices* (pp. 1–40). Springer.

Harmer, J. (2015). *The practice of English language teaching* (5th ed.). Pearson Education.

Körkkö, M., Kyrö-Ämmälä, O., & Turunen, T. (2016). Professional development through reflection in teacher education. *Teaching and Teacher Education*, *55*, 198-206. https://doi.org/10.1016/j.tate.2016.01.014

LaBoskey, V.K. (2004). The Methodology of Self-Study and Its Theoretical Underpinnings. In J. J. Loughran, M. L. Hamilton, V. K. LaBoskey, & T. Russell, (Eds.), *International Handbook of Self-Study of Teaching and Teacher Education Practices*. Springer. https://doi.org/10.1007/978-1-4020-6545-3_21

Lantolf, J. P. (2000). Introducing sociocultural theory. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 1–26). Oxford University Press.

Liu, K. (2015). Critical Reflection as a Framework for Transformative Learning in Teacher Education. *Educational Review*, *67*(2), 135–157. https://doi.org/10.1080/00131911.2013.839546

Mead, G. H. (1934). *Mind, self, and society*. University of Chicago.

Pham, T., & Saito, E. (2019). Career development of returnees: experienced constraints and navigating strategies of returnees in Vietnam [Article]. *Journal of Further and Higher Education*. https://doi.org/10.1080/0309877X.2019.1647333

Ramlal, A., & Augustin, D. S. (2020). Engaging students in reflective writing: an action research project. Educational Action Research, 28(3), 518-533. https://doi.org/10.1080/09650792.2019.1595079

Ryan, M. (2011). Improving reflective writing in higher education: A social semiotic perspective. *Teaching in Higher Education*, *16*(1), 99–111. https://doi.org/10.1080/13562517.2010.507311

Ryan, M. E., & Ryan, M. (2014). *Teaching reflective learning in higher education*. Springer.

Schön, D. A. (2017). *The reflective practitioner: How professionals think in action*. Routledge.

Wotring, A., Chen, H., & Fraser, M. (2024). They are talking, but is it productive? Exploring EFL students' small group talk. *TESOL Quarterly*, *58*(1), 251–279. https://doi.org/10.1002/tesq.3227

Conversation with Professor Constant Leung

Aek Phakiti[1]

The University of Sydney

ABSTRACT

This conversation with Professor Constant Leung explores key issues in language learning, how we understand and assess language proficiency, and the emerging roles of artificial intelligence in language education and assessment. Drawing on his extensive experience, Professor Leung shares insights into how languages are taught, learned, and evaluated, particularly in multilingual and multicultural contexts. The discussion also offers valuable reflections for language educators, researchers, students, and anyone interested in how AI is shaping the future of learning.

Keywords: artificial intelligence (AI), language learning, language proficiency, multilingualism, translanguaging

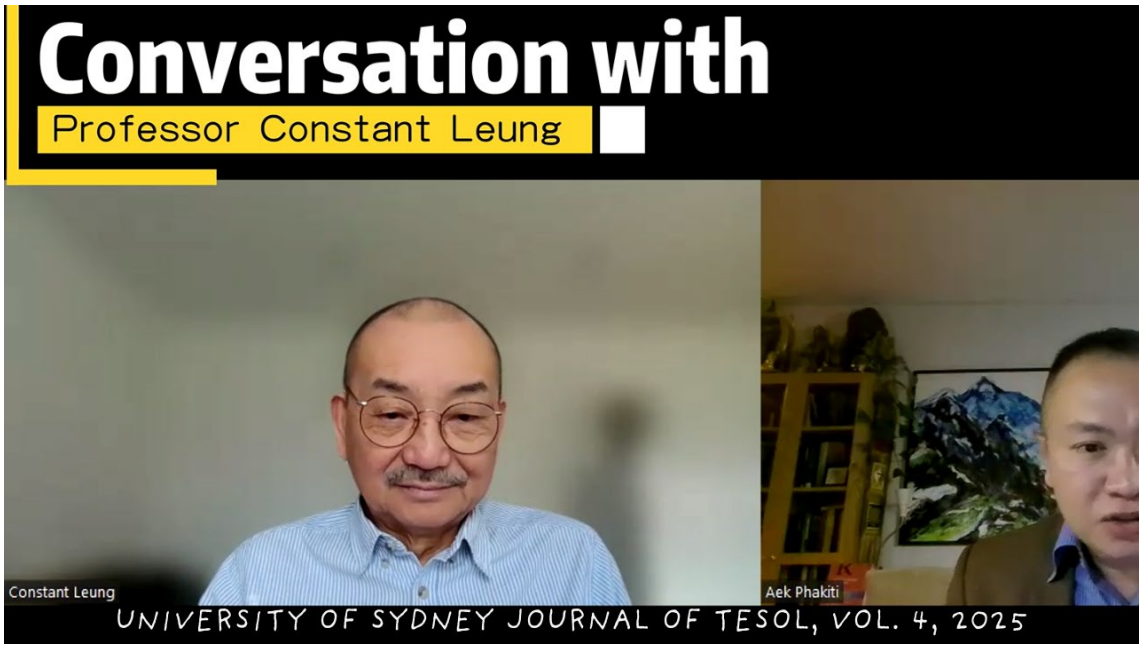
[1] Correspondence: Aek Phakiti: aek.phakiti@sydney.edu.au; http://orcid.org/0000-0002-7929-8924

ISSN 1834-4712 (Online).

DOI: 10.5281/zenodo.17591072

Citation: Phakiti, A. (2025). Conversation with Professor Constant Leung. *University of Sydney Journal of TESOL*, *4*, 179–182. https://doi.org/10.5281/zenodo.17591072

THE CONVERSATION



Note: Click the image to watch.

URL

https://www.youtube.com/watch?v=MCrfMhy8o-A

Scan Me

SUMMARY OF THE CONVERSATION

This eight-part conversation presents a rich and reflective dialogue with Professor Constant Leung, a leading scholar in applied linguistics and language education. Each episode explores a distinct thematic focus, offering insights into his academic journey, pedagogical and assessment perspectives, and critical reflections on language proficiency and assessment in the past decades.

Episode 1 (About Professor Constant Leung)

Professor Leung discusses his longstanding engagement with applied linguistics, emphasising the centrality of language in educational and social contexts. He reflects on his motivations for studying second language acquisition, language policy, and assessment. This episode outlines his teaching portfolio at King's College London, including modules on language acquisition, policy, and assessment. He highlights the evolving nature of curriculum design in response to technological and societal changes.

Episode 2 (Language Learning)

Professor Leung traces the historical development of second language acquisition (SLA) theories, noting a shift from psycholinguistic models to more socially situated understandings of SLA. He underscores the importance of integrating individual, social, and linguistic dimensions in SLA research.

Episode 3 (Language Proficiency)

Drawing from his 2022 article, Professor Leung critiques traditional notions of language proficiency. He argues for a more nuanced, context-sensitive understanding that accounts for diverse communicative practices and challenges the assumption of a fixed linguistic standard.

Episode 4 (Language and Communication)

This episode explores the communicative turn in language teaching and assessment, which is informed by the anthropology and ethnography of language use. Professor Leung examines how language use in real-world contexts complicates standardised teaching and assessment models.

Episode 5 (Model of English Language)

Professor Leung critiques the dominant model of English used in education, highlighting its narrow focus on standardised, professional varieties. He argues that this excludes everyday and culturally diverse language practices, calling for a broader, more inclusive understanding of English that reflects its global and pluralistic use.

Episode 6 (Multilingual Dimensions)

Professor Leung discusses the Common European Framework of Reference (CEFR) and its expanded focus on plurilingualism and mediation. He illustrates how multilingual interaction demands dynamic, localised judgments of communicative competence.

Episode 7 (Language Assessment Back and Forth)

Professor Leung discusses the limitations of traditional language assessment, proposing a dual approach that distinguishes between stable linguistic knowledge and dynamic interactional competence. He stresses the importance of locally responsive and context-sensitive evaluation practices that reflect real-world language use.

Episode 8 (AI, Language Use and Language Assessment)

The final episode considers the implications of artificial intelligence in language learning, teaching and assessment. Professor Leung expresses cautious optimism, stressing the importance of transparency in data sources and algorithmic design to ensure valid and equitable assessment outcomes.

Conclusion

Across the eight episodes (nearly one hour long), Professor Constant Leung offers a compelling and critical exploration of language education, bridging theory, practice, and policy. His reflections challenge conventional paradigms and invite educators, researchers, and policymakers to reconsider how language proficiency, assessment, and pedagogy are conceptualised in increasingly multilingual and technologically mediated contexts. The video makes a meaningful contribution to ongoing dialogues in applied linguistics and TESOL, encouraging a more inclusive and context-aware approach to language teaching and learning.