



Hiding in plain sight: the linguistic and behavioural challenges of second language learners in academic search

Tim Leigh and Frances Johnson

DOI: <https://doi.org/10.47989/ir31ISIC65281>

Abstract

Introduction. This study examines the influence of language proficiency on academic online searching, a critical skill in UK higher education. Language has been suggested as a factor shaping search success, particularly for second language English speakers. The investigation explores links between proficiency and search task performance.

Method. Twenty Manchester Metropolitan University students completed a screen-recorded academic search task. Participants included ten first language English (L1) and ten second language English (L2) speakers. Each was asked to locate, evaluate, and record sources for a mock academic assessment.

Analysis. Data were examined through quantitative and qualitative approaches. This involved linguistic and structural coding of search queries, analysis of search engine use, and assessment of collected sources using the CRAAP test for credibility and relevance.

Results. L2 participants demonstrated weaker linguistic control, with limited lexical range, domain mismatches, and semantic errors compared to L1 peers. They also relied less on academic sites and produced lower-scoring sources under the CRAAP criteria.

Conclusion. The study reveals clear disparities in online search between L1 and L2 students, despite shared induction and academic support. Language proficiency appears central to these differences, suggesting that applied linguistics can inform strategies to enhance search practices in higher education.

Introduction

UK higher education (HE) has experienced exponential growth in two areas in recent years. The first is internationalisation: more than 750,000 international students are enrolled in UK HE, accounting for almost 25% of the student population (HESA, 2025). The second is digitalisation: institutions continue to invest heavily in organisational, technical, and pedagogical aspects of a digital campus (Skelton, 2023). With the emergence of these trends comes potential inequality. For those students whose first language is English (L1) and have attained UK secondary and further education qualifications, many will enrol into degree programmes with English language qualifications demonstrating an ability 'to understand literary and linguistic concepts, write fluently and coherently, and demonstrate excellent analytical and critical thinking skills' (gov.uk, 2025). For learners who speak English as their second language (L2), a minimum requirement for enrolment is often a proficiency level defined as the ability to 'function in a variety of environments...with a limited range of nuance and precision' (CEFR, 2025). Potential inequity regarding digitalisation commences upon enrolment, in which HE providers do not ordinarily assess or evaluate a student's level of digital literacy, whilst training of university-led technology is neither mandatory nor part of a student's academic degree programme.

Such disparities may contribute to 'the awarding gap' (OfS, 2025): a statistically significant difference in the classification of degrees awarded to students from different demographic groups. Data consistently shows UK domiciled students as more likely to achieve an upper second-class (2:1) degree than their international peers (HESA, 2025), with literature suggesting the gap is a result of systemic UK HE issues, including pedagogical practices, academic support, and assessment design. UK universities ordinarily assess and grade learners through coursework and with predetermined marking criteria. A key criterion for institutions is the use of secondary sources. To achieve a first-class degree (70% or higher) at our university undergraduates must use deep, broad, and wide-ranging sources. From professional and personal experiences leading a range of undergraduate programmes, our small-scale empirical study aimed to interrogate whether academic online search amongst L2 learners was an added obstacle. To address this question, two research objectives were formulated: to identify the linguistic control and choices of L1 and L2 HE students in query formulation for online searching (RO1), and to identify L1 and L2 HE students' information search behaviour in online searching (RO2).

Literature review

Upon enrolment, UK HE students enters a new 'information landscape' (Lloyd, 2010, p. 9). An environment in which a learner must understand, connect, and utilise a range of social, academic, and operational networks, sources, and systems. For past generations, this ecology existed primarily on campus and developed through human interaction, but for students in the digital age 'an information ecology is often developed through online interactions' (Chang, et al. 2020, p. 3). Within university libraries, sessions and information are frequently delivered on information search strategies to navigate students through this landscape. Yet despite these provisions, much of the literature on online search suggests HE students present a lack of understanding and a range of inefficiencies in search (e.g. Georgas, 2014, Dahlen et al., 2020), with frequent examples of 'rudimentary search strategies and deficits in search behaviour' (Zlatkin-Troitschanskaia et al., 2021, p. 1997). Limitations include the use of unaccredited sources in assessment submissions (Judd and Kennedy, 2011), the inability to recognise issues with search results (Salehi et al., 2018), and the longstanding phenomenon of 'Google dependence' (Bloom and Deyrup, 2015, p. 597). Literature across both library and information studies and applied linguistics has sought to analyse and recommend support for international students, centring on library instruction, English language tuition, and digital skills training. Details on the specific content of library workshops are limited in the literature, however studies such as Sapa (2024) identify common themes across institutions. These tend to include the development of students' information literacy, the use of source

evaluation tools, and the importance of academic integrity. English language tuition for international students is often optional for students, with attendance excluded from a student's academic progression. Digital skills training appears the least consistent of the three areas of support, with universities exhibiting differences in the quantity and mode of training provision.

Yet, decades of search as learning research describes a critical and fundamental context in which the student searches for information to learn about a topic. As identified by Marchionini (1989) such searching would be recognised as a part of exploratory search. Three decades ago, Kuhlthau's (1991) modelling of the exploratory stage recognised the students' somewhat unfocused articulation of their query as driving a need to explore and interact with information retrieved. More recently Urgo and Arguello (2022) describe searching to learn as an iterative process with the searcher acquiring new knowledge by interacting with multiple sources, assessing, and judging the information. Further recognising the cognitive processing in learning searches, they refer to constructivism theory of learning to describe the student searcher as an 'active participant in their own learning process' (Urgo and Arguello, 2025, p. 372). As such, search as learning research aims to characterise search behaviour in learning environments and to explore the factors that may impact on individual's search activities (such as query formulation). For example, the impact of the searchers' expert domain knowledge on search activities (such as number of queries, length of query, pages visited) may be compared as in Zhang et al. (2015). In this study, we explore the possibility that the language attainment of L2 students may be a factor impacting search activities during learning. Drawing on studies within applied linguistics that posit the deep, often invisible academic literacies, such as expected argumentation structures in written assignments, that can be taken for granted but which many L2 students may be challenged to understand (e.g. Baker-Bell, 2020). It may be that L2 students face similar conscious and unconscious challenges when required to search for active learning, and as such, may be 'hiding in plain sight' (Margolis et al. 2002, p. 6) from staff and institutions. Identifying the ways in which language proficiency impacts students' search as learning presents an opportunity to understand the challenges faced and the language needed for students' learning to search.

Method

Data was collected from students on a foundation year BA Honours Business Management undergraduate degree programme, who volunteered to participate in a screen-recorded academic search task. The task rubric asked participants (ten L1 and ten L2) to search and collate relevant academic sources for a draft summative assessment related to the role of Snapchat in business marketing and advertising. To address RO1, a structural-linguistic analysis of participants' search terms was conducted via a computer-based linguistic programme WordSmith Tools (Scott, 2008). This software was able to identify and organise each participant's search language into frequency, cluster length, and part of speech, allowing query (re)formulations to be analysed, coded, and duplicated into a search process flowchart. To investigate RO2, screen recordings were analysed for search engine preference and participant source collation. Analysis commenced with the identification and coding of academic and non-academic sites, with anonymised collated sources then graded by two academics using the CRAAP test (Blakeslee, 2004). By reworking information literacy standards into five criteria the test provides a checklist to evaluate source currency, relevance, authority, accuracy and purpose, and each source was allocated a score (0 (poor), 1 (Fair), 2 (Good), 3 (Excellent)) for each criterion and combined to give an overall score. Although not universally applied across UK HE institutions, this evaluative tool has been regarded as a useful quantifier for students to evaluate sources (e.g. Lewis, 2018) and has been used in similar studies focused on L1 and L2 university students (e.g. Sendurur and Gabriel, 2023). Data triangulation from RO1 and RO2 enabled the development of a holistic search process for each participant, presenting search stages, query formulations, search engine use, and recorded sources (see Figure 1). This representation of each search session enabled analysis of the development of search language, its impact on search outcomes, and comparison between L1 and L2 participants. It also provided an

opportunity to identify recurrent, rhetorical linguistic and behavioural actions, a lens commonly adopted in applied linguistics, specifically within digital genre analysis (e.g. Swales, 1990, Hafner and Pun, 2020).

Participant X	Search move 1	Search move 2	Search move 3	Search move 4	Search move 5
Search language	Is snapchat good for business?	Snapchat business	Snapchat business UK	Social media UK business	Social media marking business
Search engine	Google	Google Scholar	Google Scholar	Google Scholar	University Library
Recorded source CRAAP score	No source	No source	4	8	8

Figure 1. Example of a participant search process.

Findings

In response to RO1, a structural-linguistic analysis of search terms identified differences between L1 and L2 participants' linguistic choice and control. Data from L1 students' search (example shown in Figure 2 below), included the use of the interrogative form. Coded as detrimental to the search task (QRF-), question forms accounted for almost one fifth of L1 search language. Beyond this structure, much of the L1 search language was coded as beneficial to the task (QRF+), with the effective use of hypernyms, fixed expressions, and synonyms related to the search task. Conversely, L2 query formulations revealed a limited range of lexis and a multitude of detrimental linguistic choices, including domain mismatches, lexical divergence, and semantic incompatibility. Figure 3 (below) shows an example L2 participant's search language, with search moves coded as detrimental (QRF-) to successful task completion. In Figure 3, for example, the L2 participant misplaces snapchat as the subject (rather than the object) of the query in search moves 1 and 3 due to the misuse of the preposition 'for'.

Search stage	Search move 1	Search move 2	Search move 3	Search move 4
L1 participant	Is snapchat an effective tool for marking and advertising	Social media marketing campaign	Endorser definition	Snapchat for marketing disadvantages
QRF code	QRF-	QRF+	QRF+	QRF-

Figure 2. Example L1 participant's search language.

Search stage	Search move 1	Search move 2	Search move 3	Search move 4
L2 participant	Points of good marketing for snapchat	Snapchat response to online marketing	Marketing for snapchat	Snapchat news stories
QRF code	QRF-	QRF-	QRF-	QRF-

Figure 3. Example of L2 participant's search language.

Regarding RO2, two key findings emerged from the analysis of search behaviour. First, search engine preference differed between groups. Most L1 participants began on Google, with seven later moving to the academic sites of the university library, Google Scholar, and Science Direct. Their overall visits were balanced across academic (9) and non-academic (9) sites (see Table 1). In contrast, L2 participants commenced searches on a wider range of engines, with Google used in only 4 of 10 first visits and the university library being their sole academic resource. L2 participants also shifted to non-academic sites more often. Whilst only one L1 participant moved to such a site after the first search, four L2 participants did so, turning to Google and Yahoo in later visits (see Table 2). Overall, L2 participants visited academic sites less frequently and relied more heavily on non-academic sources than their L1 peers.

Search engine	Search engine category	Search visit 1	Search visit 2	Search visit 3	Overall visits	Academic vs Non-academic visits
Google	Non-academic search engines	8	0	1	9	9
Microsoft Edge		0	0	0	0	
Yahoo		0	0	0	0	
University Library	Academic search engines	2	4	0	6	9
Science direct		0	1	0	1	
Google scholar		0	2	0	2	

Table 1. L1 search engine visits (academic vs non-academic search engine).

Search engine	Search engine category	Search visit 1	Search visit 2	Search visit 3	Overall visits	Academic vs Non-academic visits
Google	Non-academic search engines	4	2	1	7	10
Microsoft Edge		2	0	0	2	
Yahoo		0	1	0	1	
University Library	Academic search engines	4	2	1	7	7
Science direct		0	0	0	0	
Google scholar		0	0	0	0	

Table 2. L2 search engine visits (academic vs non-academic search engine).

The second key finding concerns search efficacy and task performance. Comparative analysis of CRAAP results highlights clear differences in performance. L1 students demonstrated stronger search skills and collated more appropriate sources. Seventeen of their sources scored 8 or above (defined as good to use and cite), compared with only six for L2 participants. The average CRAAP score further reflects this gap: L1 participants achieved 9.5 (good to cite), whilst L2 participants averaged 5.2 (adequate but not recommended for citation). An additional factor was the distribution of low-quality sources. Amongst L1 participants, 7 of 24 sources (29%) fell into the 3–4 CRAAP score range. By contrast, 70% of L2 sources scored at this low level, including 30 graded as very questionable, with a CRAAP score of 3.

Conclusion

This study highlights how disparities in academic search practices exist between L1 and L2 UK HE students. Whilst a range of institutional support mechanisms are available, including library workshops, English language classes, and digital skills sessions, these interventions are typically generic in content and optional in attendance. The case of a single student cohort, all of whom are enrolled onto the same degree programme, provides a telling example. Despite equal exposure to university introductory support and provision, significant variation emerged in the students' ability to identify and retrieve deep, broad and wide-ranging sources to meet university assessment criteria. These findings suggest academic online search may form part of a hidden curriculum within UK HE. A widely cited definition of this phenomenon is the 'unstated norms, values and beliefs transmitted to students' through formal curricular content and classroom social relations' (Giroux and Penna, 1979, p. 22). In practice, research has translated this definition into a number of HE contexts, including assessment feedback (Johnson, 2020), student disabilities (Van Hees, et al. 2018), and the student profile (Speirs, 2020). In this study, institutional level information and guidance related to finding and evaluating online sources can be found on the university's library website, and explicit reference to the use of sources is intergraded into the institution's assessment marking criteria. Nonetheless, this study implies the disparate practices of online academic student search are potentially hidden from academics, particularly with reference to linguistic control and choice. This raises some concern that university guidance may not be effective in addressing the hidden curriculum. Library guides typically encourage the student to consider their topic and identify key terms. Use of Boolean operators may be explained to encourage thinking about how to express the query. The findings from this study suggest linguistic choices and control of the L2 UK students impacted on search efficacy. Although library instruction seems to suggest that a level of language proficiency and flexibility is necessary to identify the key terms and their relationships within a search.

Future exploration of the data aims to investigate how applied linguistics, specifically digital genre analysis, can contribute to search as learning research. To develop broader academic literacies amongst L2 UK HE students, a frequently adopted approach within applied linguistics is genre-based pedagogy. Described as a practice providing 'students explicit and systematic explanations of the ways language functions in [specific] contexts' (Hyland, 2003, p.18), the integration of a linguistic intervention could play a positive role in search as learning. Grounded in the principle that language is shaped by the communicative purposes of specific discourse communities, genre-based pedagogy provides a structured approach to developing students' awareness of genre and often follows a three-stage teaching and learning cycle: deconstruction, joint construction, and active/independent construction. These stages are pivotal in demystifying the conventions of specific genres and making implicit knowledge explicit for novice learners (e.g. Fenwick and Herrington, 2022). In this study, it was evident L2 participants were limited in their ability to be active in their own learning as they lacked not just the language, but the metalanguage of search. Teaching search as a genre aims to address this challenge by learning from clear explanations how language use achieves effective searches and demonstrating that search can be learnt. Learners analysing search are encouraged to think about the information need and how it is expressed as a query. Further analysis of differences in the queries would help learners recognise reasons for these and the impact on the results. Furthermore, learning to search involves assessing and learning from the retrieved information. In conclusion, this study highlights the need to practise search meta/language to minimise the risk of learners hiding in plain sight. The proposed approach to teach search as a genre strives to reveal search can be learnt and practiced for students to become competent in their search as learning journey.

About the authors

Tim Leigh is a Senior Lecturer in the School of English at Manchester Metropolitan University, UK. His research interests include English for academic purposes, internationalisation, and inclusive education. He can be contacted at t.leigh@mmu.ac.uk

Frances Johnson is a Senior Lecturer in the School of English at Manchester Metropolitan University, UK. She has a PhD in computational linguistics from the University of Manchester and her current research interests include information retrieval and behaviour. She can be contacted at f.johnson@mmu.ac.uk

References

- Baker-Bell, A. (2020). *Linguistic justice: Black language, literacy, identity, and pedagogy*. Routledge.
- Blakeslee, S. (2004). The CRAAP test. *LOEX Quarterly*, 31(3), 4.
- Bloom, B., & Deyrup, M. M. (2015). The SHU research logs: Student online search behaviors transcribed. *The Journal of Academic Librarianship*, 41(5), 593–601. <http://dx.doi.org/10.1016/j.acalib.2015.07.002>
- Chang, S., Gomes, C., & McKay, D. (2020). The digital information ecology of international students. In C. Gomes & S. Chang (Eds.), *Digital experiences of international students: Challenging assumptions and rethinking engagement* (pp. 1–20). Routledge.
- Council of Europe. (2024). *Common European framework of reference for languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing.
- Dahlen, S. P., Haeger, H., Hanson, K., & Montellano, M. (2020). Almost in the wild: Student search behaviors when librarians aren't looking. *The Journal of Academic Librarianship*, 46(1), 102096. <https://doi.org/10.1016/j.acalib.2019.102096>
- Department for Education. (2024). *GCSE subject-level guidance for English language (2022)*. Gov.uk. <https://www.gov.uk/government/publications/gcse-subject-level-conditions-for-2022/gcse-subject-level-guidance-for-english-language-2022>
- Fenwick, L., & Herrington, M. (2022). Teacher use of genre pedagogy: Engaging students in dialogue about content area language during text deconstruction. *Language and Education*, 36(1), 43–58. <https://doi.org/10.1080/09500782.2021.1912082>
- Georgas, H. (2014). Google vs. the library (part II): Student search patterns and behaviors when using Google and a federated search tool. *portal: Libraries and the Academy*, 14(4), 503–532. <https://doi.org/10.1353/pla.2014.0034>
- Giroux, H. A., & Penna, A. N. (1979). Social education in the classroom: The dynamics of the hidden curriculum. *Theory & Research in Social Education*, 7(1), 21–42.
- Higher Education Statistics Agency. (2025). *Higher education student data: Alternative providers*. HESA. <https://www.hesa.ac.uk/data-and-analysis>
- Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of Second Language Writing*, 12(1), 17–29. [https://doi.org/10.1016/S1060-3743\(02\)00124-8](https://doi.org/10.1016/S1060-3743(02)00124-8)
- Johnson, I. (2020). What hides beneath? An evidence-based take on the hidden curriculum of assessment feedback. In T. Hinchcliffe (Ed.), *The Hidden Curriculum of Higher Education* (pp. 5–28). Advance HE.

- Judd, T., & Kennedy, G. (2011). Expediency-based practice? Medical students' reliance on Google and Wikipedia for biomedical inquiries. *British Journal of Educational Technology*, 42(2), 351–360. <https://doi.org/10.1111/j.1467-8535.2009.01019.x>
- Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42(5), 361–371.
- Lewis, A. B. (2018). Tips from the experts: What does bad information look like? Using the CRAAP Test for evaluating substandard resources. *Issues in Science and Technology Librarianship*, (88). <https://doi.org/10.5062/F41N7ZC4>
- Lloyd, A. (2010). *Information literacy landscapes: Information literacy in education, workplace and everyday contexts*. Elsevier.
- Marchionini, G. (1989). Electronic problem scenarios: Integrating computers into instruction. *Journal of Education for Library and Information Science*, 30(3), 165–176.
- Margolis, E., Soldatenko, M., Acker, S., & Gair, M. (2002). Peekaboo: Hiding and outing the curriculum. In E. Margolis (Ed.), *The hidden curriculum in higher education* (pp. 11–30). Routledge.
- Office for Students. (2023). Office for Students. <https://www.officeforstudents.org.uk>
- Salehi, S., Du, J. T., & Ashman, H. (2018). Use of web search engines and personalisation in information searching for educational purposes. *Information Research*, 23(2). <http://informationr.net/ir/23-2/paper789.html>
- Sapa, R. (2024). Thematic scope of workshops offered by the libraries of the top universities. *The Journal of Academic Librarianship*, 50(3), 102730. <https://doi.org/10.1016/j.acalib.2024.102880>
- Sendurur, E., & Gabriel, S. (2024). Evaluation of search engine results pages in various languages and contents: The changes in strategies and criteria. *The Electronic Library*, 42(2), 173–192. <https://doi.org/10.1108/EL-05-2023-0111>
- Scott, M. (2008). Developing wordsmith. *International Journal of English Studies*, 8(1), 95–106.
- Skelton, N. (2023). *Digital strategies in UK higher education: Making digital mainstream*. JISC. <https://www.jisc.ac.uk/reports/digital-strategies-in-uk-higher-education-making-digital-mainstream>
- Speirs, N. M. (2020). The hidden curriculum as doxa: Experiences of the working class. In T. Hinchcliffe (Ed.), *The hidden curriculum of higher education* (pp. 130–141). Advance HE.
- Swales, J. M. (1990). *Genre analysis*. Cambridge University Press.
- Urgo, K., & Arguello, J. (2022a). Learning assessments in search-as-learning: A survey of prior work and opportunities for future research. *Information Processing & Management*, 59(2), 102821. <https://doi.org/10.1016/j.ipm.2021.102821>
- Urgo, K. & Arguello, J. (2022b). Understanding the pathway towards a searcher's learning objective. *ACM Transactions on Information Systems*. 40(4) 1–43. <https://doi.org/10.1145/3495222>
- Urgo, K. & Arguello, J. (2025). Search as learning. *Foundations and Trends® in Information Retrieval*, 19(4) 365–556. <https://doi.org/10.1561/15000000084>

- Van Hees, V., Roeyers, H., & De Mol, J. (2018). Students with autism spectrum disorder and their parents in the transition into higher education: Impact on dynamics in the parent-child relationship. *Journal of Autism and Developmental Disorders*, 48(10), 3296–3310. <https://doi.org/10.1007/s10803-018-3593-y>
- Zhang, X., Ding, X. Y., Wang, G. S., & Ma, L. (2020). Investigating the influences of social overload and task complexity on user engagement decrease. *Total Quality Management & Business Excellence*, 31(15–16), 1774–1787. <https://doi.org/10.1080/14783363.2018.1509698>
- Zlatkin-Troitschanskaia, O., Hartig, J., Goldhammer, F., & Krstev, J. (2021). Students' online information use and learning progress in higher education: A critical literature review. *Studies in Higher Education*, 46(10), 1996–2021. <https://doi.org/10.1080/03075079.2021.1953336>

© [CC-BY-NC 4.0](#) The Author(s). For more information, see our [Open Access Policy](#).