

This is a repository copy of *Balancing evidence-informed language policy and pragmatic considerations:Lessons from the MFL GCSE reforms in England*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/217249/>

Version: Published Version

Article:

Marsden, Emma orcid.org/0000-0003-4086-5765 and Hawkes, Rachel (2024) Balancing evidence-informed language policy and pragmatic considerations:Lessons from the MFL GCSE reforms in England. *Language, Society, and Policy*. ISSN 2515-3854

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Balancing evidence-informed language policy and pragmatic considerations: Lessons from the MFL GCSE reforms in England

Emma Marsden and Rachel Hawkes | 17th December 2024 | Policy Papers

- Policy change:
 - New curricula for GCSE French, German, and Spanish were released by the [DfE in 2022](#), for first examination in 2026. About 250,000-300,000 16-year-olds are likely to take these exams every year.
 - Many aspects remained the same as [DfE \(2014\)](#), including similar rationales to broaden horizons and promote skills including meaningful interaction.
 - Innovations included: reduced amounts of grammar; detailed specification of sound-writing relations; detailed specification of word patterns; tests of core literacy; a word list of 1250/1750 items (Foundation/Higher), of which 85% are high-frequency, from which reading and listening exams must be created; recognition that irregular forms are learnt holistically; explicit assessment of inferencing skills; opportunity to broaden language use to cultural/historical/geographical/social/political domains; foregrounding of unprepared speech, emphasising comprehensibility over accuracy.
- Research suggests that:
 - The previous, 'guide' word lists optionally provided by the awarding organisations had not been used in principled ways in exams, so the lexical content of curricula (textbooks, schemes of work) has been poorly aligned with high-stakes exams.
 - New frequency-informed GCSE vocabulary lists provide better preparation for many different types of text, including adolescent fiction and A-level exams.
 - The highest-frequency words substantially overlap across different corpora.

- A series of four exams have generally included about 1350/1750 unique words, suggesting the newly specified number of words allows awarding organisations to produce appropriately different exams year-on-year.
- Knowledge of high-frequency vocabulary is positively associated with inferencing skills and with self-efficacy.
- We argue for:
 - Closer scrutiny of how the government's subject content is operationalised in exams and criteria.
 - Better alignment between the pace of policy change and the pace of dissemination of peer-reviewed research.
 - More opportunities for educators and policymakers to engage in research (often kept behind paywalls).
 - More, high-quality, applied linguistics research conducted in schools about languages other than English.
 - Research into desirable balances between teaching (and assessing) language itself and the understanding of other societies and cultures.

Introduction

Context and problem. In England, for approximately 80% of students in primary and secondary schools, French, German, and Spanish are the only languages learnt, other than English. Every year, over half a million 14-16-year-olds follow a GCSE curriculum in these languages. After about 400-450 hours of instruction in secondary school, 15-16-year-olds take the GCSE exam, which (whether one likes it or not) heavily shapes curricula, materials, and pedagogy. Despite the high stakes, the *language* content of these exams had been largely overlooked by researchers and policymakers. This is surprising, given the strong relationships between assessment, curriculum, pedagogy, and motivation. Some researchers had attempted to promote motivation and/or strategies to help students access unfamiliar language that could be encountered during the course and exams; but MFL GCSE uptake was continuing to decline (perhaps plateauing in 2024).

Policy initiative. Between 2019 and 2024, we contributed to a DfE review of the GCSEs in French, German, and Spanish and its subsequent operationalisation in accredited exam material. The review aimed to make these languages more accessible to more students via changes that wouldn't require large investment in training or modifying whole school policy. As implementation of these changes by Ofqual and the awarding

organisations runs on, it has become clear that the [Overton window for change](#) was relatively—and perhaps necessarily— narrow.

Aims of this policy paper. Part 1 summarises the GCSE revisions, focusing on changes relating to vocabulary. Part 2 addresses misconceptions—among some commentators and stakeholders during the consultation and to date—and discusses relevant research evidence. In Part 3, we consider lessons about language assessment policy change.

Part 1: Changes to the GCSE in French, German, and Spanish

Key similarities and differences. In 2022, after a consultation period (typically fairly rapid as with most consultations, and also postponed due to the pandemic), the [DfE \(2023\)](#) released a new subject content. That document underlies GCSEs for examination from 2026 onwards and shapes most secondary teaching materials. Key rationales for studying languages (communication, cultural enrichment, further study) remained, and one rationale was added: “to better understand relationships between the foreign language and the English language” ([DfE, 2023, p. 3](#)). The previously specified personal and transactional topics ([DfE, 2014](#)) were changed to: “a range of broad themes and topics which have, for example, cultural, geographical, political, contemporary, historical, or employment-related relevance” ([DfE, 2023, p. 3](#)). This was partly in response to ongoing concerns that curricula heavy with personal-social topics [were alienating students](#) ([NALA, 2020](#)) within certain socioeconomic/sociocultural demographics. Another innovation was the explicit listing of sound-writing correspondences, indicating core literacy content, and two related (semi-)realistic tests (reading aloud and writing down speech). Substantial parts of grammar (e.g., French subjunctive) were removed, and previous unnecessary conflation between grammar and the lexicon were resolved. The new annexes provided greater *detail* about the remaining grammar, which some commentators (e.g., some journalists and vocal language educators) misinterpreted as a bigger emphasis on grammar. In fact, grammar that can now be assessed is substantially reduced, thus better (though not perfectly) aligning with evidence about the slow rate of language development. Principles of parsimony and usefulness informed some decisions, e.g., prioritising the simpler and more frequent periphrastic future (*ils vont dormir*) over the inflectional future (*ils dormiront*). The main assessment tasks remained similar, though with an intended reduced focus on ‘general conversation’ in favour of shorter, *unprepared* interactions around stimuli (short text, pictures).

Changes to the vocabulary. Perhaps the most fundamental change was vocabulary. GCSE exams must be created using the 1,250/1,750 (Foundation/Higher) lexical items listed by the awarding organisations; off-list language can be rewarded, but full marks must be achievable using the listed words. There were misunderstandings among commentators (including some influential educators and journalists) that using wordlists was out of kilter with other jurisdictions, but several other education systems (increasingly) use obligatory, frequency-informed core word lists (reviewed by [Finlayson, Marsden, and Hawkes, 2024 \[OASIS Summary\]](#); [Marsden et al., 2023 \[OASIS Summary\]](#)) Other innovations included:

1. 85% of 1200/1700 of the items must be from among the 2,000 most frequent lemmas (headword + inflected forms) found in a very large multi-genre spoken and written corpus, with the other 15%, plus 50 multiword phrases and cultural/geographical terms, from any frequency band;
2. intentional assessment of lexical inferencing skills for off-list words;
3. glossing for some off-list words;
4. a definition of cognates (for reading assessments only, as orthographic cognateness is more transparent than phonological cognateness);
5. acknowledgement that many highly irregular forms tend to be stored holistically and so deserve 'curriculum space' as listed 'words';
6. reduction of listed word patterns, leaving a research-informed selection of derivations for reading only, based on frequency, productivity, form complexity, and reliability of meaning (see [DfE, 2023](#); [Finlayson et al., 2024 \[OASIS summary\]](#)).

Part 2: Misconceptions: Policy Change and Research

During and following the GCSE review and consultation, high-quality research and its dissemination tried to keep pace with an appetite for speedy policy change. Although tentative findings related to vocabulary in GCSEs emerged during the review, *international researcher peer review* was needed for their wider use. This lag between policy change and rigorous evidence probably fuelled controversy and debates characterised by false dichotomies, some of which we address here.

Misconception 1: "1,250/1,750 words are not enough".

During the review, awarding organisations aired concerns that lists of 1,250/1,750 lexical items would not provide enough words to create exams, though they could not report how many lemmas had been needed for exam creation to date, perhaps because the technology had not been available in languages other than English. Using the tool [MultilingProfiler](#), developed parallel to the review, we found that each exam used about [600 \(foundation\) and 780 \(higher\) lemmas](#); and to create four different exams, about [1,350 \(foundation\) and 1,750 \(higher\) lemmas](#) were needed ([Dudley & Marsden, 2024](#)). This suggests that the number of words prescribed by the new policy could easily support exam creation.

Publicly available research now also spotlights the need to change the existing optional lists, which have been developed since 1988 using subjective topic-driven selection principles to support teaching and materials. [Marsden et al. \(2023\)](#) found that (i) the length of these current lists varied substantially between awarding organisations and languages (for example, Edexcel had on average about a third more lemmas in their lists than AQA); (ii) the lists had been used sparingly in the exams – an average of 47% of the words had *never* been used in four series of listening and reading GCSE exams; (iii) the lists provided insufficient coverage (around 70%) of the exams, too little for sufficient comprehension; (iv) a significant proportion—approximately 25%—of the lexical items used in these exams were *not* on the word lists.

Misconception 2: "High frequency words are less useful; low frequency words are more useful".

Teachers had reported that topic-based language was demotivating students (e.g., [NALA, 2020](#)). In part to address such concerns, the new frequency-informed word selection principle aimed to "enable material relating to most broad themes and topics to be used, and [...] unlock a wide range of spoken and written texts." ([DfE, 2023, p. 3](#)). However, awarding organisations, groups, and individuals expressed grave concerns that using 'frequency' would not support our students. The concern is illustrated by a lesson observation undertaken by a representative from a cultural institute who asked a student "How do you say, '*I would like a single ticket to x*'?" – a highly-specific transactional phrase requiring a low-frequency term. The student not knowing this term was used by the visitor as evidence of inappropriate language teaching. Yet, the student *was* easily able to communicate this same need with '*I want to go to x by train*'.

Now, evidence shows that frequency-informed lists provide much better coverage of even the *current* GCSE texts (see [Finlayson et al., 2024 \[OASIS summary\]](#); [Marsden et al., 2023 \[OASIS summary\]](#)). Additionally, a new GCSE list better covers [a range of different relevant texts](#): adolescent fiction, language from the internet, A level exams, and GCSE exams. Crucially, this better coverage is due to the principled selection of *content* words, (not function words, as some claimed). Moreover, the awarding organisations' own word selections for the new GCSE overlap with each other substantially, indicating an intuitive core word set. Although *mathematically* their lists could have differed by approximately 900 words (resulting in only 49% overlap between them), there is in fact a very high degree of overlap (73%-94%) between three independent interpretations (by AQA, Edexcel, LDP/Eduqas) of the list creation parameters.

We also now know that, year on year in *every exam*, only [a very small number of words were being used](#)—just 200 on average. Such a small number of words nevertheless covered about 80% of all words used: a very narrow pool constituting most of the language assessed. On the other hand, 11-13% of the words had only ever appeared once across four exam series, with two-thirds of those being low frequency, making it *highly unlikely* that students, in our limited exposure context, would know or infer them (low levels of inferencing are found [even among high achieving students](#)). The finding that 87% of running *words* (tokens) were high frequency supports the stipulation that '85% of a word list must be high frequency'. But critically, the new policy now requires awarding organisations to sample from a *larger* pool of high-frequency lemmas year on year.

Misconception 3: "The suggested language corpora are not relevant".

Concerns were voiced in the media, a petition, webinars, and blogs that the corpora (texts) used to provide frequency data were not relevant to our students (though, as far as we are aware, no viable alternatives were suggested). Now, published evidence shows that enough of the same lemmas are high frequency *across several relevant* corpora. For example, between 1,250 and 1,640 of the most frequent 2,000 words in the *exemplar* corpora by [DfE \(2023\)](#) (i.e., the corpora used to create the Routledge frequency lists and used by awarding organisations for the GCSE lists) are [shared across three other corpora of adolescent-relevant language](#): internet language; GCSE and A level exams; and adolescent fiction. Relatedly, it's important to recall that, in fact, an average of only about 892 (Foundation) or 1,265 (Higher) different lemmas

need to be selected from among the most frequent 2,000 lemmas—a reasonably sized pool from which to select a common core, thus easily avoiding words perceived to be overly 'corpus-specific'. Also, using *general* (rather than specialist) corpora is arguably appropriate for the *General Certificate of Secondary Education*, given the difficulty of predicting any specialist future language needs of *all* 16-year-olds.

To accommodate variation in language over time, we also recommended including in the subject content (or Ofqual's regulations) some allowance for new/updated corpora to be used and/or for some of the 15% of words from any frequency band to be replaceable, at pre-specified intervals of time (to reduce unnecessary disruption to teaching and materials). We do not know whether such principles will be implemented, but, in any case, it is possible—and important—for students to personalise their vocabulary for *production* (speaking and writing) at any point in time.

Misconception 4: "Specifying language content will lead to dry pedagogy and stifle proficiency"

The press, social media, webinars, and publications alluded to unevidenced conflation between a greater specification of language and an assumed, inevitable dry pedagogy resulting in poorer (i) ability to use the language (proficiency), (ii) inferencing (working out the meaning of unfamiliar words), and/or (iii) cultural knowledge.

'*Poor proficiency*'. Ample evidence already supported the idea that language components (e.g., vocabulary, grammar) [predict and drive proficiency](#) (Jeon & In'nami, 2022). And we now have evidence from our own context that better knowledge of vocabulary, grammar, and phonics (very) strongly predicts proficiency scores, including GCSE grades, the international DELF/DELE, and communicative tasks such as unprepared writing, role plays, and unprepared conversations (see [Dudley et al., 2024 \[OASIS summary\]](#); [Dudley & Marsden, forthcoming](#)).

'*Cultural content reduced*'. There were (ongoing) concerns that specifying language content *necessarily* detracts from cultural content. Yet, as Woore et al., (2022, p. 149) wrote: "The National Centre for Excellence for Language Pedagogy has shown that it is possible to integrate cultural elements into a scheme of work which is carefully sequenced in terms of grammatical structures and high-frequency vocabulary (as documented in their '[Cultural collection](#)')". See further [examples for Key Stage 4](#).

Similarly, it is possible to create exam content that adopts cultural themes for the new GCSE, though whether this will be realised is yet to be seen.

'Inferencing skills won't develop'. Another concern was that using vocabulary lists will curb the development of inferencing. In fact, the new GCSE explicitly assesses inferencing, and a [curriculum with meticulously defined core language content can still promote inferencing skills](#). Also, research on vocabulary sizes indicates that most learners will [not know all words on the lists](#) (and see [OASIS summary](#)); so unfamiliar language will still require inferencing skills. Moreover, better knowledge (breadth and depth) of the most frequent 2,000 words is strongly positively associated with (a) more accurate inferencing, (b) greater self-efficacy (confidence when faced with challenge), and (c) more interest in the text (see e.g., [Dudley & Marsden, forthcoming](#)).

Summary. Reform to the lexicon used in assessments was needed. Given that to understand a text, around 90-98% of its words need to be known—through existing knowledge or inferencing—, the high proportion of unfamiliar vocabulary in exams (and associated teaching materials) to date could be *one* cause of demotivation. It could also have contributed to the observed 'severe grading' of languages: other things being equal, students have received lower grades in languages than other subjects. Put simply, if curriculum content doesn't *reliably* appear in an exam, and yet other, *unexpected* content does appear, then grades are likely to be out-of-kilter with school subjects that have closer alignment between curriculum and assessment.

Part 3: Challenges in Operationalising Language Assessment Policy

Defining tests to assess a body of knowledge is not straightforward. Here we discuss some concerns about the operationalisation of the new GCSE policy.

Good tests of core literacy? The new subject content lists sound-writing relations to promote core literacy, which supports reading-aloud and transcribing/notetaking, and other aspects such as vocabulary learning. Designing such tests is complex, and we mention here two concerns. First, the policy's focus on assessing phonics almost exclusively with words that are on the wordlist raises concerns about whether the assessment will validly measure generalisable knowledge of sound-writing relations (or, simply, the spelling of known words). Second, it was deemed not possible for one task to assess both reading-aloud and comprehension of that text, as Ofqual were

concerned that this would threaten test validity. As a result, reading aloud and the subsequent conversation can effectively be largely independent, potentially weakening links made—during instruction and assessment—between literacy and comprehension. The counterargument is that a valid assessment of language should in fact assess *both* ‘form’ (such as knowledge of sound-spelling relations) and ‘meaning’ (comprehension).

Assessing culture and intercultural competence? The new subject content sought to promote a wider range of intrinsically interesting themes, replacing specific transactional/personal/social topics with “broad themes or topics with relevance to the counties or communities where the language is spoken. These could cover, for example, cultural, geographical, political, contemporary, historical or employment-related aspects” (DfE, 2023, p. 5). However, some of the sample exams and specifications, now accredited by Ofqual, continue to foreground everyday personal transactional situations (see, for example, Edexcel’s role plays).

The British Academy’s response to the GCSE consultation presented interesting ideas about assessing culture. Why were these suggestions seemingly beyond the potential for acceptable change (the apparent [Overton window](#))? A major issue may be that we need a better understanding of *how* to validly and reliably assess cultural competencies for *all* learners in this context, if we want to avoid (a) resorting to symbolic tokenism (‘name three important buildings in Paris’), or (b) invoking performative rote-learning of prefabricated language (‘give your opinion in 50 words on cultural artefact/extract x’). We need a community consensus on situating ‘*language*’ learning in this context in relation to other knowledge areas like history, art, geography, literature, music, and (social) sciences. Additionally, research must explore different methods of assessing cultural understanding and intercultural competence. Crucially for a *languages-for-all* policy, these methods must account for the exceptionally wide range of learner differences and varied conceptions of ‘culturally engaging’ content. Such research must also work with a realistically learnable *amount* of language, a convenient proxy for which is likely to be receptive knowledge of [between 550 and 1650 words](#) for most 16-year-olds, given current curriculum time allocations.

Genuine tests of communicative speech? The new subject content specifies ‘unprepared conversation/interaction’, replacing the previous ‘speak *spontaneously*’ (‘spontaneity’

being difficult to assess). It also requires only '*clear and comprehensible*' speech, aiming to reduce pressure at this stage to produce accurate and complex speech, and to decrease washback into rote-memorisation of long chunks of prefabricated speech. Curiously, some high-profile educators and assessors have dismissed such rote-preparation for speaking exams. [But we now have evidence that 84% of year 11 students \(from 2022 and 2023 cohorts\) reported having 'learned off by heart' answers for their speaking exam.](#) 69% that they had been asked to do so by their teacher, and 57% reported also using some of these answers in their writing exam. This demonstrates the use of rehearsed language as a short-cut to fluent, accurate, and complex language. We are therefore concerned about how the subject content's requirements relating to speech have been operationalised. Relatedly, genuine interaction involving speaking *and* listening is not always convincingly operationalised: the role plays' cues are presented in sequence, which, along with the preparation time given, removes any *real* need to listen to the interlocutor (examiner), even though officially the skill of listening *is* assessed by the role play for Edexcel.

The opportunity was there to set interactive tasks and develop criteria that genuinely promote unprepared speech and reward communication, comprehensibility, and content, reducing the (explicit or implicit) reliance on indices of accuracy and complexity. Such indices—implicit in task design, rubrics, and marking scales for decades—have in part driven a rote-learning of phrases that many learners cannot manipulate to create genuine meaning.

Lessons on the Nature of Evidence-informed Policy Change

Our semi-centralised examination system. In England, we have a government-controlled curriculum and yet *commercially competitive* awarding organizations. This competition probably provoked disparities in how the subject content was operationalised in exams, driving decisions that undermined the spirit of some policy changes. We call for debate about whether this hybrid system serves the best interests of children's learning and motivation. *If* we agree that curricula should be determined by government (arguably reasonable, in a democracy with free schooling), then perhaps that same government should also be directly responsible for *how* that content is assessed. Ofqual, as current regulator, provides only very high-level oversight of how the subject content is interpreted and operated by different bodies; for example, they do not directly monitor the sampling of language content year on year. Furthermore,

the regulation of how policy is interpreted into assessment practice seems largely inscrutable once policy has been approved by the DfE.

Transparency and equity in operationalising subject content Greater transparency and collaboration would help to ensure equitable and effective assessment. We hope that future collaborations between teachers, researchers, and materials and test developers will be at least as (if not more) open and direct as our experiences with the awarding organisations. For example, one concrete collaboration has been the [MultilingProfiler](#), which now allows all stakeholders to check exam and teaching texts against various word lists.

Pace and mechanisms for change. The pace of policy change outstripped the rate at which research could be validated. Ideally, policy innovators could commission internationally excellent, double-blind peer-reviewed published evidence *before* implementing changes. In the GCSE review, with more time, more specialists could have contributed. We successfully requested that non-panel linguists review the subject content and appendices; but more data, available sooner, would have informed debates, potentially reducing feelings of exclusion or fear among some stakeholders.

Research into language education. Most *high-quality* research in instructed language contexts has been about the teaching of *English* to *self-selecting* learners in *higher* education. We call for rigorous *empirical* research in primary and secondary schools in majority Anglophone settings on:

- Effective balances between, on the one hand, teaching the most useful language (vocabulary, grammar, core literacy) as rapidly as possible and, on the other hand, meaningful activities that engender intrinsic interest (i.e., we need more alternatives to supplement the approach described by [Marsden & Hawkes](#), but similarly detailed);
- The effects of these GCSE policy changes. Many primary and secondary educators are now creating schemes of work to systematically revisit core language in rich and varied contexts. However, data on the effects of these changes on *proficiency* and *motivation* are needed. We have a sizeable dataset from 2022-23 prior to the changes (see [ComLaP](#)), but this needs replicating in 2026-27.

Assessment policy changes alone are no silver bullet. Other innovative policies and practices obviously deserve serious long-term evaluation. We echo calls for research into approaches such as: diversifying the languages, including better formal recognition of our multilingual population; immersion approaches; language- and culture-awareness programmes; assessment criteria that better align with evidence about instructed language development; and viable alternatives to an exams-driven system, investigating the fundamental idea of measurement-driven instruction.

We hope that our work will stimulate more curiosity-driven interactions at research-policy-practice interfaces. We need greater support for a 'research-intrigued' culture among more teachers and policymakers. But critically, we need more higher education researchers who are informed about the *realities* of practice and *constraints* of policy change.

References

Dudley, Amber, and Emma Marsden. 2024. 'The Lexical Content of High-Stakes National Exams in French, German, and Spanish in England', *Foreign Language Annals*, 57: 311–38. <https://doi.org/10.1111/flan.12751>

Dudley, Amber, and Emma Marsden. Forthcoming. *Language Proficiency and Its Components: The Case of Adolescent Beginner to Low-Intermediate Learners in Schools in England* (Multilingual Matters).
<https://sites.google.com/york.ac.uk/comlap/> [accessed 16 December 2024]

Dudley, Amber, Emma Marsden, and Giulia Bovolenta. 2024. 'A Context-Aligned Two Thousand Test: Towards Estimating High-Frequency French Vocabulary Knowledge for Beginner-to-Low Intermediate Proficiency Adolescent Learners in England', *Language Testing*. <https://doi.org/10.1177/02655322241261415>

Finlayson, Nathalie, Emma Marsden, and Rachel Hawkes. 2024. 'Creating and Evaluating New Vocabulary Lists for Adolescent, Beginner-to-Low-Intermediate Learners of French, German, and Spanish', *Language Teaching Research*.
<https://doi.org/10.1177/13621688241288877>

Jeon, Eun Hee, and Yo In'nami (eds). 2022. *Understanding L2 Proficiency: Theoretical and Meta-Analytic Investigations* (John Benjamins)

Marsden, Emma, and Rachel Hawkes. 2023. 'Situating Practice in Curriculum Design in School Foreign Language Education', in *Practice and Automatization in Second Language Research: Perspectives from Skill Acquisition Theory and Cognitive Psychology*, ed. by Yuichi Suzuki (Routledge), pp. 89–118

Marsden, Emma, Amber Dudley, and Rachel Hawkes. 2023. 'Use of Word Lists in a High-Stakes, Low-Exposure Context', *The Modern Language Journal*, 107.3: 669–92.
<https://doi.org/10.1111/modl.12866>

Cite this article

Marsden, Emma and Rachel Hawkes 2024. 'Balancing evidence-informed language policy and pragmatic considerations: Lessons from the MFL GCSE reforms in England'. *Languages, Society and Policy*. <https://www.lspjournal.com/post/balancing-evidence-informed-language-policy-and-pragmatic-considerations-lessons-from-the-mfl-gcse>

About the authors

Emma Marsden is professor of second language education at the University of York, having started her career as a school languages teacher. She directs two open research repositories: iris-database.org and oasis-database.org. She was Associate Editor then Journal Editor for *Language Learning* (2015-2022) and directed the National Centre for Excellence for Language Pedagogy (2018-2023; ldpedagogy.org).

Rachel Hawkes is Director of Languages and International Education for The Cam Academy Trust, which is Oak National Academy's curriculum partner for languages at KS2, KS3 and KS4. She was previously Head of Modern Languages in two secondary schools, AST, then Senior Leader, and co-directed the National Centre for Excellence for Language Pedagogy (2018-2023).