UNIVERSITY *of* York

This is a repository copy of *A narrative account of experiences at research-policy-practice interfaces in England—warts and all.*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/231372/

Version: Accepted Version

Article:

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

A narrative account of experiences at research-policy-practice interfaces in England—warts and all.

AIMS OF THE ARTICLE

We provide a reflective, narrative account of work at research-policy-practice interfaces in foreign education in England; a public record of work undertaken between 2018 and 2023 for a very large government Department for Education (DfE) funded initiative to support pedagogy, curricula, and assessment for French, German and Spanish, via professional development (PD) and resources. The DfE's overarching goal was to improve foreign language pedagogy for 11–16-year-olds and (thus) increase the numbers of students studying a language GSCE at age 14. To that end, we were tasked with developing, sequencing, and implementing a large-scale, cohesive set of research-informed PD events and resources for language teachers. The PD and resources were to support the approaches to pedagogies, curricula, and assessments recommended by a policy-endorsed review (Teaching Schools Council, [TSC] 2016). In this article, we offer an account of and reflections on our thought processes, decisions, and activities at key stages of the project. We capture several collaborative dimensions, including the relationships between: the authors; the authors and their wider team; the authors and the DfE; and the relationships between the initiative itself (authors and/or DfE) and schools and teachers. We describe clusters of interrelated activities, driven at pace, many of which were determined by pragmatic constraints, and end with frank discussions about challenges and legacies. In doing so, this piece may serve as a reflection space for others interested in engaging in large-scale research-policy-practice opportunities. We aspire to provide "evidence which fuels the hope that strong collaborative endeavours between traditional academic researchers, teachers and practitioner researchers are possible" (Consoli & Dikilias, 2021: 350).

THE AUTHORS' PARTNERSHIP

We believe that our researcher-practitioner partnership is key, and we give our credentials here, as recommended by Rose & McKinley (2022). Hawkes self-identifies as a practitioner first and foremost and she has always been very keen to keep her affiliation with her school. She is a highly experienced primary and secondary school teacher of Spanish, German, and French, a senior school leader and advisor, an author of widely used school textbooks and curricula, and a prolific professional development provider, with over 6,000 twitter followers. She has a PhD on spontaneous talk in the classroom, has sat on several

policy committees, and has recently been a named author, due to her conceptual, practical, and leadership roles, on academic publications with Marsden. Marsden, a qualified teacher of Spanish, French, and English taught in secondary schools 1994-98 (and English overseas from primary to adults for short periods) and has self-identified as a university-based researcher and HE teacher for over two decades. She has led many academic publications and research projects on fundamental and applied research on learning and teaching in classroom and laboratory settings, and was an associate and Journal Editor of *Language Learning* (2015-2022), which mainly publishes research with a strong theoretical focus with a smaller proportion related to instruction. Like Hawkes, she has contributed to policy panels and advisory committees.

The history of our relationship is worth describing briefly, as it may illustrate how such collaborations can emerge. Having been aware of each other's work for some time, we initially met face to face in about 2012 on an arms-length policy-advisory group. In about 2014, with others, we unsuccessfully applied for research council funding to explore ways of better assessing oral communicative proficiency. Our work together on the panel of nine for the Review of MFL Pedagogy (TSC, 2016) marked the beginning of the close collaboration documented in this article.

Our partnership was critical for *obtaining* the DfE funding for NCELP, as it added unique strength to our response to the government tender. It was also critical for *sustaining* the funding (achieving three extensions to the DfE contract, and a subsequent large contract to Rachel for creating more resources building on NCELP's work). Rachel brought to the partnership her knowledge of teacher experiences, needs, and appetites, her insight of and empathy for school constraints and affordances, her reputation in the school community, a track record of embedding phonics, classroom target language, and culture in language teaching, and, perhaps most precious, a curiosity and openness about research. Emma brought knowledge of a wide range of learning theories and empirical evidence from education and applied linguistics, a capacity to discuss and identify their potential practical relevance drawing on her experience in context-relevant settings, and her ability to synthesize and communicate complex ideas and findings from international peer-reviewed research.

Together, we worked with resource developers and CPD providers to weave these different affordances into our work with and for teachers. However, we acknowledged throughout and still today, a difference in our discourse traditions and communities of practice: Rachel had greater expertise of oral, social media, and grey literature (policy documents) discourses, and Emma of written and professional academic discourses. We

believe that a profound mutual respect for both these communities and praxes was essential for our work. This respect emerged naturally from the start, but was also 'learned' through honest dialogue, occasionally challenging for us. 'Power relations' (an issue often discussed in academic writing on research-practice interfaces) were almost always unintrusive, harmonious, and productive, though they were necessarily discussed explicitly in a conversation at the outset and once or twice subsequently. Funding had been awarded to the University only, which had ultimate responsibility for delivery; and so it was mutually agreed that, in cases of any disagreement (of which there were thankfully very few), Emma would have to make the final decision and, in daily communication with Rachel, she took ultimate responsibility for hiring, line management, and strategy in terms of key aims, deliverables, pace of work. She, in discussion with Rachel, led on ensuring our work was research-informed where possible, but—where research was absent or weak in terms of relevance or clarity (a relatively common situation)—our explicitly discussed and agreed stance was that practitioner expertise must come first. As the stages described below progressed, we perceived a slight shift in 'power relations' with co-leadership morphing into more leadership from Rachel in terms of content, pace, and direction, in close discussion with and support from Emma and the University. We hope that our account illustrates that a "research-practice relationship can be equitable, effective, and mutually beneficial" (Sato, 2023, p. 1).

PART 1: BACKGROUND TO THE CONTEXT AND PROJECT

1.1 Context.

In England, the key foreign languages (FL) taught in schools continue to be western European languages—specifically, French, German, and Spanish—in part due to their role in colonial empire-building. Whether rightly (these languages are useful) or wrongly (underpinned by legacy injustices and low investment), these three languages usually constitute the *only* significant experience of languages other than English for approximately 80% of children across primary and secondary schools. Education in these languages constitutes the main livelihoods of approximately 14,500 secondary school teachers and affects the livelihoods of perhaps 33,000 primary school teachers[1].

FL study is compulsory between ages 8 and 14, and almost half (approximately ¼ million) of all 14-year-olds stop after only approximately 250 hours of lessons. Most of the

---

[1] assuming two teachers teach some language at each of the 16,800 primary schools

rest drop languages after about 450 hours, after taking the GCSE (the General Certificate of Secondary Education) as part of a suite of high-stakes national exams taken at age 16. Most of this instruction takes place during secondary school given the limited provision of FLs in primary schools and, generally, poor continuity between primary and secondary education (Graham et al., 2016). Falling numbers studying FLs from aged 14 onwards mean fewer graduates with advanced inter-cultural and linguistic skills—and therefore teachers—leaving reported personal, social, economic, and cultural deficits across multiple societal domains. Underlying this situation are several macro socio-political and socio-linguistic factors that cannot be influenced by most practising educators. However, teachers and materials designers *can* influence an albeit small number of variables—some of which were within NCELP's remit and are the focus of the current article.

1.2 Current curriculum and pedagogy and political investment to support change

Since the 1980s, language education in England has adopted a largely topic-driven curriculum, in part influenced by (mis- or over-) interpretations of communicative approaches. Educators, textbook publishers, and awarding organizations have selected topic- and function-specific sets of words and phrases, such as the weather, food and drink, free-time, daily routine, travel, opinions. These topics determine language selection, such that, for example, past tense is taught mainly during the topic of holidays; future tense is taught mainly during the topic about future plans. Heavily topic-driven sequencing is evidenced in the optional word lists published by the three commercial awarding organizations[2] as guides for teachers' planning and publishers' textbooks, exerting a powerful washback on high-selling materials (e.g., Hawkes & Lillington, 2016).

Key concerns with this approach relate to the high risk that some vocabulary and grammar are experienced only once or twice, in 'topic-silos' or fixed phrases, and practised 'in' specific skills (listening, reading, writing, speaking), rather than systematically established and practised *across* different semantic and linguistic contexts, in both receptive and productive modes, and in both oral and written modalities.

Dissatisfaction with school FL education can be inferred from a number of sources over decades: low student uptake (Churchward, 2019; Joint Council for Qualifications, 2024), teacher dissatisfaction with topics (NALA, 2021), perceived (Coffey, 2016) and actual

---

[2] The Assessment and Qualifications Alliance (AQA, 2016), the English branch of the Welsh Joint Education Committee (Eduqas, 2019), and Pearson Edexcel (2018). Note, changes are in place for curricula starting in 2024, for GCSE examination in 2024).

(Ofqual, 2019) difficulty of foreign languages at GCSE, lack of alignment between curriculum and assessment (Dudley & Marsden, 2024), widespread rote-learning for spoken and written production (Dudley & Marsden, under contract), slot and fill phrase-based learning (Mitchell, 2002) promoting (slow) creative use among only small proportions of learners (Myles et al., 1998), uninspiring classroom activities (Wingate, 2016), concerning commentaries from school inspectorates (Ofsted, 2008, 2011, 2021), and, more indirectly, also perhaps reflected in teacher shortage (Colleen & Duff, 2024) and weak engagement with research findings (Marsden & Kasprowicz, 2017).

Calls to increase the quantity and quality of languages education have been heard for decades. The latest central government policy (albeit under the previous government) aims to have 90% of students take a language GCSE by 2027. As part of working towards this target, we were invited to take part in the "Review of MFL Pedagogy" in 2015-2016, led by the TSC (2016), commissioned by the DfE. This publication (so-called 'grey literature') detailed 15 recommendations about pedagogy, curricula, assessment, delivery, and provision. The recommendations related to pedagogy and curriculum largely followed a componential view of development and proficiency, whereby components of knowledge (e.g., of words, phrases, grammar, sound-spelling relations) are established in different modes and modalities, and automatised to support skill acquisition. To support teachers with those specific pedagogy and curriculum recommendations, the government issued a competitive tender for a pilot for a "National Centre for Excellence for Language Pedagogy" (NCELP) to lead, co-ordinate, and monitor the work of nine 'Lead Schools' each with 4 'Hub' secondary schools (a network of 45 secondary schools in total). The contract was awarded to us, held by the University of York, in December 2018, and over 4.25 years this pilot initiative received a total of £5.2m from the DfE, which we supplemented with additional funds[3]. Critically, this total was split over four *separate* contracts of two years, one year, one year, and then three months. Whilst this may seem a trivial detail, it posed serious challenges for implementing research-informed change on a large-scale in a short time. Retention of colleagues, teacher learning, community cohesion, and evidence gathering all require longevity. Nevertheless, by contract end in March 2023, we had a team of about 30 resource developers, PD providers, researchers, administrators, and 12 in-school teachers delivering PD, working locally and nationally.

---

[3] From the ESRC IAA, The University of York, Research England, and the Higher Education Innovation Fund.

PART 2: FOUR RELATED AND OVERLAPPING PHASES OF WORK

The pilot involved multi-faceted and iterative approaches to research-informed PD (Watson et al., 2021) and resources (see https://ldpedagogy.york.ac.uk/resources/; Finlayson & Marsden, 2022; Marsden & Hawkes, 2023). Marsden et al. (2023) constituted the final report for the government funder. Here, we artificially— for the sake of coherence—divide our work into four 'phases'. However, it is critical to appreciate that our process was emergent, iterative, and dynamic, reacting to heavily contextualised events, suggestions, requests, and needs from policymakers, teachers, commentators, and society (the pandemic).

2.1 Phase One (years 1-2): Identifying aims, selecting research, extracting principles, establishing models of working

2.1.1 Establishing aims and scope. As the project was directly using government funds (taxes), we were not as free as we might have liked to determine aims, methods, or evaluations. A key argument that persuaded the government's Treasury to release the funds to the Department for Education to support FL education mainly consisted of increasing the numbers taking FL GCSE, particularly in socio-economically deprived areas. This in turn became a target for us: to increase uptake of, specifically, French, German, and Spanish in the network of 45 secondary schools by 15%, via research-informed innovation to curriculum and pedagogy in Key Stages 3 and 4 (ages 11-16), as per recommendations of TSC (2016). The terms 'excellence' and 'pedagogy' in the name of the centre were chosen by the DfE[4]. Prior to signing the contract, we told DfE colleagues that a sustainable increase of this size was highly unlikely as many factors influence GCSE choices, including global sociolinguistic phenomena, teacher supply, effects of school performance league tables, schools' policies on compulsory versus optional language GCSE, students' socio-economic background etc. Nevertheless, each party signed, despite potential financial penalties to the university if the target was not met[5].

2.1.2 Establishing working relations. The DfE selected the nine Lead schools prior to our own contract signing, for expediency. Their selection criteria related to existing excellent FL

---

[4] Throughout, we use pedagogy to refer to the 'what, when, how' of teaching, and thus include curriculum design and sequencing, classroom techniques, and specific ideas and resources that support practitioners.
[5] Indeed, this target was not fully met by the end of the contract; see discussion by Marsden et al. (2023: 34-37).

GCSE uptake and achievement. However, this order of events, although certainly accelerating the start of the Centre's work, caused initial misunderstandings, because schools and the Centre had different understandings of the aims and methods of the programme. For example, some teachers were not aware that they themselves would be undertaking research-informed PD on the innovations detailed by TSC (2016).

A contrasting approach to teacher development would have been to invite teachers to (1) identify their challenges in terms of student outcomes, (2) recognise what they know already, and (3) articulate what they would like to know to improve student outcomes (Mujis et al., 2014; Timperley, 2011). However, our funding was to address challenges identified *a priori*, and to follow a given model of PD to deliver content to support the implementation of components of TSC (2016). The situation we faced is captured by Mujis et al. (2014: 247-248):

> "The problem with this […] as a starting point is that the need to know something new is identified by someone external to the group of teachers (e.g., a policy official or a researcher) without the participating teachers necessarily understanding the reason why it is important to know it or being committed to doing so. Under these circumstances, the goals belong to others who are taking responsibility for promoting the professional learning. Teachers then choose whether to engage or to resist."

In our case, engagement could perhaps have been enhanced and resistance reduced if the DfE's criteria for teachers wishing to join the programme had included an interest in developing identities as *professional learners* (Mujis et al., 2014). This may have reduced any dissonance caused by misunderstandings that we detected (subjectively, implicitly) in some early interactions during which the formal contracts between schools and NCELP were established.

On the whole, this dissonance was soon resolved, though it is hard to pinpoint the reasons with confidence. Some resolution was reached by us—and particularly Rachel as a trusted and respected practitioner—visiting individual schools prior to the first PD residential. Nevertheless, two or three of the 18 teachers at that initial event still demonstrated overt reluctance to engage in new research-informed ideas. One of these teachers left the project and was replaced by another. Any remaining sense of unease that we had perceived gradually resolved during subsequent PD events, perhaps as intrinsic interest, trust, and familiarity grew. This demonstrated to us the importance of sustaining PD to support innovation over relatively long periods (more than several weeks or months). Interestingly, one of the teachers who had shown perhaps the greatest initial reluctance became one of the most

enthusiastic advocates for the programme having adapted the ideas and materials to her own context.

Another dynamic, context-bound factor that makes it difficult to identify why dissonance resolved was that the GCSE review started in year 2 of the project and this promised to better align curriculum content, research-informed ideas, and high-stakes assessments. Although the GCSE review was interrupted by the pandemic, it is likely that the prospect of this better alignment and of being at the forefront of such change facilitated assonance; teachers could then perhaps more readily accommodate (rather than superficially assimilate) ideas from research within their own frameworks of operation.

There was one exception to this general trend: one of the nine Lead School worked with us for three years but perhaps never recovered from this initial dissonance and did not continue on in year four to join the others offering CPD to wider audiences. Their large school academy trust (i.e., collection of schools) had pervasive top-down practices for classwork, homework, and technology that the lead teachers found could not easily work with new approaches to pedagogy or curricula. This was not surprising, as it reflected the powerful agency of these large school trusts. But follow-up data would be very useful to explore the longer-term effects of having engaged in the PD.

2.1.3 Identifying key domains and research-informed principles. Between 2015-2023 (and continuing to date), the two authors had many informal (prior to contract) and formal (during contract) conversations, by email, over zoom and at the kitchen table, about key ideas in language education. Our discussions were, in part, seeded by TSC (2016): curriculum design, elements of pedagogy, the nature of proficiency and its components, assessment regimes and types. Throughout, and particularly in initial months, one strong characteristic of our interactions was that we asked each other many questions that led to rich dialogue. Rachel was probably more likely to ask Emma open questions in initial months (though the reverse also happened); the questions sometimes emerged following an interaction with a practitioner or from reading. Rachel, already having completed a PhD, consumed and synthesised research findings at pace and with criticality. We both drew on syntheses and on our other collaborators (Graham, Woore, Kasprowicz, and later Emma informally sought advice from several international vocabulary experts). We carefully considered a *range* of potential theories and pedagogical frameworks (including immersion, task-based teaching and learning [TBLT], content and language integrated learning [CLIL]), and we both had insights into evaluating the importance of their *relevance to the context*. Emma recalls some questions

being some of the most challenging and meaningful in her professional experience to date. She found that her concurrent role as journal editor for *Language Learning*, a different space in many respects, was complementary as it provided a breadth of cutting-edge ideas and references. Some questions were, necessarily, very wide-ranging, and a joke we began to share was: "This is probably the topic for five PhD theses, but … what do you think about [x]". Emma was also invited to comment on 'education influencers' in England—both language educators and generalists—whose work is not normally within scope of peer-reviewed applied linguistics. These exchanges illustrated to us two oft-observed challenges to working at research-practice interfaces: (i) how the laser-focus of much academic research means that it can often not adequately address contextualised needs, and (ii) the need for managing expectations, among *all* parties, about what research can offer (Lightbown, 1985, still relevant 40 years on).

It was over a period of about a year and a half that several fundamental decisions were taken. In considering which curricula and pedagogies would be included in our PD programme, we discussed:

(i) the likely acceptability of ideas and the feasibility of their operationalisations within the Hubs and beyond;

(ii) innovations that hadn't yet found their way into previous research-informed initiatives;

(iii) the quality of research evidence to date and its potential for *context-relevance*;

(iv) the alignment with the TSC (2016), and in particular, aspects of that review that were likely to be the most challenging (i.e., most unlike widespread practice);

(v) our funder's Key Performance Indicators (KPIs);

(vi) the potential for conveying clear messages in the time available. It became apparent that certain approaches (e.g., strong forms of CLIL or TBLT) were not feasible within budget and timeframe given the amount of training and change that would be needed. They were also deemed riskier than others in terms of their potential to improve achievement and/or motivation for all students, given a lack of rigorous evidence that is of relevance to *this* context.

Although some of these pragmatic rationales would lead to potentially controversial decisions, they necessarily informed our PD. Next, we provide more detail on five dimensions of this scene-setting work.

(i) *Identifying context-relevant learning theories:* We jointly alighted upon skill acquisition theory for several reasons: the explicit nature of learning in schools, the skills-

based test regime, the usefulness of 'practice' as a framework for developing curricula and activities (Suzuki, 2023), and growing bodies of evidence supporting a cognitive, componential view of proficiency (e.g., Andringa et al., 2012 and comprehensive meta-analyses such as Jeon & In-ami, 2022; Zhang & Zhang, 2022). We thus included PD themes on: the nature of practice (skill acquisition and cognitive interactionist theories); vocabulary (the role of frequency; effective interventions); phonics (sound-writing relations, core literacy); and grammar (input-based focus on form/s; the limited role of explicit information; pros and cons of deductive versus inductive approaches; trapping forms in 'task-essential' or 'task-useful' production activities; meaningful practice); the (constrained) role of corrective feedback types.

(ii) *Identifying motivation-related theories*: For this context, appealing to extrinsic, integrative, or instrumental motivations is not reliably supported (e.g., Taylor & Marsden, 2014) as most teenagers have relatively little exposure to societies where the languages are spoken, and there is little convincing and/or reliable motivation from others (e.g., employers/universities/parents/media to inspire an 'ought to L2 self'). We, of course, wished to improve these situations via promoting international interactions and cultural encounters, but these were already promoted and prioritised by national cultural institutes and existing professional networks (and was the focus of one of our practitioner-delivered PD session). However, it was acknowledged that the most likely motivators were *intrinsic*, such as enjoyment, sense of achievement, and engagement with the subject as experienced in school. Thus, the PD and our wider work foregrounded the notion that students feel more motivated when they (feel like they) are achieving more (Garon–Carrier et al., 2016), linked to self-confidence in a capacity to deal with challenge (self-efficacy).

(iii) *Defining an amount for a realistic curriculum for all*: There was surprisingly little research evidence available about (components of) proficiency attainable after five years of study in secondary school. We felt this perhaps constituted a disappointing report card for applied linguists in England in terms of addressing questions relevant to context (Ortega, 2013). Tentative evidence gave us indications about amount of vocabulary that could be learned (now reviewed by Dudley & Marsden, 2024; Dudley et al., 2024; Marsden & Hawkes, 2023; Marsden et al., 2023) and we combined this with practitioner expertise and previous GCSE subject contents (DfE, 2022/2015) to work out 'what' and 'how much' language might be reasonably covered in the time available for most teachers and students.

(iv) *Envisaging a sequenced curriculum:* With theories identified and amounts of knowledge estimated, we then had to create a curriculum that established, revisited, and

integrated knowledge (including vocabulary, grammar, and phonics), and wove this around culturally rich content that builds towards the ability to communicate in different modes and modalities. This required exceptionally meticulous planning and creativity, to promote progression via both intentional learning (i.e., planned, explicit) and incidental learning (i.e., both implicit and self-induced explicit).

(v) *Connecting new ideas alongside teachers' existing practice:* As Mujis et al. (2014: 247-248) note: "learning […] occurs by making patterns that connect *existing* knowledge to new knowledge […]. It makes sense, therefore, that when building new professional knowledge and refining skills, teachers are assisted to make these connections so they can understand *what is the same and what is different* about the kind of thinking and practice being promoted" [italics ours]. Supporting teachers in making these connections involved various steps. For example, Rachel analysed a popular textbook to illustrate various problems with heavily topic- and phrase-driven approaches, such as: vocabulary was not revisited (see also Laufer, 2023 for similar findings); low frequency vocabulary had a significant presence in specific topic-oriented conversations; vocabulary selection did not lend itself to teaching regular grammar or phonics patterns. We were also asked by the DfE to create two documents that laid out points of alignment between our work and (a) the Early Career Framework (DfE, 2019) and (b) new school inspection framework (Ofsted, 2023). Whether or not we agreed with these policies, they were shaping educators' work and so it was responsible to situate our work against those frameworks. Other ways of connecting new knowledge to existing knowledge included: creating observation schedules spotlighting potentially innovative pedagogy; comparing our word lists with existing word lists; illustrating differences between the novel input-based activities that induce *form-meaning* connections versus those that focus on form or on meaning alone; providing questions to steer teachers' own analyses of their existing schemes of work in light of the novel ideas.

2.1.4 Selection of and communications around research. Intermediaries working at research-practice interfaces usually synthesise and select research findings to communicate key ideas in a short time; but any selection brings subjectivity. The research selected had to support recommendations from TSC (2016) and be high-quality. We had to foreground potential relevance to our contexts, where (i) exposure is limited, (ii) learning is mainly intentional and explicit, and (iii) motivation (of any type) cannot be assumed. Also, we did not want to use these public funds to re-hash ideas that had been mainstream for years, such as heavily topic sentence-based 'slot and fill' approaches—after all, it was on that watch that

numbers of students were tumbling. Nor did we need to duplicate the broad-church support, general networking, and information-sharing offered by the national professional association, the Association for Language Learning.

We also wanted to nurture research literacy, to promote questioning and awareness of the limitations of research. We strived to do this throughout the PD, providing (i) research-informed ideas and findings as stimuli for considering their relevance to practice, as well as (ii) concrete examples of classroom resources to illustrate how the ideas might influence pedagogy, materials, and assessment. How we did that and who undertook different roles emerged as the initiative progressed, described briefly here. However, the division of labour we depict is not intended to suggest clearly delineated roles or interactions—'blurred lines' were apparent at every interface much of the time.

OASIS summaries (oasis-database.org; Alferink & Marsden, 2023) were selected mainly by the HE researchers, the doctoral and post-doctoral resource developers, and, though less often, Rachel and the teacher PD providers. These summaries were provided as readings prior to specific PD sessions and/or linked within PD material as suggestions for follow-up reading. The summaries provided stimuli for discussion and illustrated some of the underpinning evidence (an approach also recommended by Graham, 2024; see Marsden & Dudley, in progress, for how summaries were used and the effects they had). The selection of summaries was always read by Rachel to check for comprehensibility and relevance, which occasionally led us to revise the summaries for greater clarity. We collaborated with researchers on specific themes: Suzanne Graham (University of Reading), on rich texts, strategies, and self-efficacy; Rowena Kasprowicz (University of Reading), on input-based grammar teaching, grammar assessment, primary-secondary transition, and an app 'Gaming Grammar' (Kasprowicz, et al., 2019); Robert Woore (University of Oxford) on phonics (i.e., decoding and encoding literacy), speaking assessments, and vocabulary learning and teaching; Natalie Finlayson (University of York) on selecting, defining, and testing vocabulary. We also worked with other kinds of experts (René Koglebauer, a teacher educator at the University of Newcastle; and David Shanks, a leader in a school cluster in London) on school-based approaches to increasing uptake of language study and on technology in the classroom. Marsden led on creating short synthesis and rationale documents that condensed key research, drew out principles, highlighted complexities, and

noted areas where research could not offer clear implications[6]. The themes of these documents included vocabulary and grammar teaching and learning; meaningful practice; assessment of componential knowledge such as vocabulary, grammar, phonics (given that holistic assessment of skills was already better established in practice). Hawkes was responsible for achieving relevance and accessibility for teachers.

2.2 Phase Two (years 1-3): PD model and events

2.2.1 Model of PD.

Our model of PD was largely prescribed by the DfE, in terms of its frequency, amount, length, agents and interactants. Whilst it is not known which (if any) theoretical model of teacher development underpinned the DfE's approach, many characteristics seemed broadly in line with what is known about effective PD in terms of having positive effects on teachers' attitudes, practice, and learner outcomes (e.g, Cordingley et al., 2015; Desimone, 2009; Education Endowment Foundation, 2021; Kennedy, 2016; Lynch et al., 2019; Pringle et al., 2020; Taylor et al., 2015). The PD programme was long (over two-four years); provided *subject-specific* focus; offered opportunities to develop professional identity, autonomy, and judgment; allowed new ideas to be contextualised and embedded over time; was co-developed and co-taught by combinations of researchers, (expert) teachers, and teacher educators*;* provided research-informed pedagogical materials that could be adapted in response to students' needs*;* facilitated communities of learning in which role models demonstrated expertise in enacting change themselves. The model of PD in years 1-3 is presented in stylized format in Figure 1.

Insert Figure 1 about here.

2.2.2 Materials development

To illustrate key ideas, we made dozens of examples of adaptable classroom materials freely available on a database, a 'Resources Portal' (now held at

https://ldpedagogy.york.ac.uk/resources), for teachers to edit for their own use. Also, we endeavoured to engage the specialist teachers in creating materials themselves in a resources

---

[6] Illustrating this challenge at research-practice interfaces, one academic was unable to describe ideas or produce resources related to motivation that were of a sufficient quality and quantity to persuade teachers that they would be used in schools. This colleague had to step down from the project within a few months.

creation project, lasting about three months. Each teacher was assigned one kind of activity to create, for specific language features and vocabulary sets, ending with a small group 'show and tell' at the second residential. NCELP's resource developers then worked on the teachers' drafts e.g., providing images, audio, and additional or revised examples. These co-produced resources were then made available to all on the Resource Portal.

However, many of the 18 teachers informally reported not having sufficient time, and/or understanding, and most (probably all) did not want to continue working on materials development. They did, however, want more resources to be made available. This situation was not in line with some recent research suggesting that materials development can be a highly productive avenue for PD (e.g., Shu et al., 2024). However, we took these concerns seriously, as described next.

2.3 Phase Three (years 2-4): Illustrating a fully resourced curriculum
Teachers requested an off-the-shelf prototype that illustrated a fully resourced curriculum with language-driven practice schedules that (a) ensured key features (vocabulary, phonics, grammar) are sufficiently revisited; (b) provided ample opportunities to listen, read, speak, and write using these features across different linguistic and semantic contexts; and (c) wove themes (based around culture, functions, notions, or interactions) *around* this language core. These requests for resources challenged our researcher-practitioner partnership. As a researcher, Emma was anxious about providing materials that might be *perceived* as promoting a homogenous, researcher-proposed 'one-size-fits all' solution. Her preference was to produce editable examples, with searchable metadata to facilitate teachers finding resources to fit into their practice. On the other hand, Rachel's greater experience of teaching, PD, and materials creation drove her preference to provide week-by-week lessons with all activities, including homework, provided in each PowerPoint. The DfE colleagues on our board supported this view. Research also suggests that resourcing at the level of full lesson guides can support teachers' learning, providing stimuli for understanding and professional growth (Reeves, 2010) and have beneficial effects when combined with PD (Piper, et al, 2018; Taylor et al., 2015). It became clear that the extent of potential novelty—i.e., moving to a language-driven curriculum with cultural content—imposed a duty on us to provide at least one full instantiation. Emma therefore supported the move, given that these fully editable materials were adaptable to context, presented with the clear caveat that they constituted illustrative examples only, and each lesson resource would be individually searchable on a database with key terms (though the latter proved time-consuming to sustain).

Accommodating more resource creation led to difficult decisions. One of these was shifting funds away from three NCELP colleagues (one academic, one teacher educator, and one school-based senior leader). Another was asking the two remaining external HE researcher collaborators to undertake more hands-on involvement in resource creation and delivering research-informed PD in schools. These shifts, albeit uncomfortable to manage, were nevertheless critical as it enabled us to hire resource developers in years three and four. These developers had modern language degrees and a range of teaching experiences, and about half also had MAs or PhDs in either literature/culture or applied linguistics. They formed a creative and meticulous team, pivotal for any endeavour to respond to teacher requests for an illustrative fully resourced curriculum and resources. However, it was challenging to recruit, train, and retain such specialists, as our funding only permitted contracts that were often shorter than one year.

So, in co-production between researchers and practitioners, we undertook the first, to the best of our knowledge, attempt at creating _fully resourced_ (with all lesson materials) practice schedules for 450 hours of instruction, with about 30-45 minutes homework per week, over five years, in French, German, and Spanish (Marsden & Hawkes, 2023). Without compromising the aims of promoting intrinsic motivation and a sense of achievement, the language was meticulously selected and revisited so that vocabulary, grammar, and sound-writing relations appeared across different linguistic and semantic contexts, without inadvertently neglecting certain words or features. By the end of the DfE contract, the entire 450-hour curricula and practice schedules had been produced, with full resources for about 250 of those hours.

We had to prioritise certain aspects of resource creation; there was insufficient time and money to create all dimensions of a holistic pedagogy, for which teachers themselves are best placed to synthesise and deliver in context. We therefore prioritised the illustration of aspects of TSC (2016) that we thought were the most _innovative_ and/or _difficult_ for teachers to undertake themselves, including:

(1) estimating how much lexical content could be learnt by most learners;

(2) selecting which lexical items are likely to be most useful using subjective teacher judgements and objective frequency-informed measures. For this, we used DfE and other funds to create a new lexical profiling tool (https://multilingprofiler.net/; Finlayson et al., 2023), as there were none available in Spanish or German, and the existing French tool (lextutor.ca) could not accommodate our purposes;

(3) selecting and sequencing vocabulary sets that (i) accommodate the introduction of new grammar and phonics, *and* (ii) facilitate revisiting that grammar and phonics, *and* (iii) supports functional, interactional, and/or cultural themes;

(4) creating input-based activities that orient learners' attention on understanding the meaning/function of grammar (the 'referential' component of processing instruction; Marsden, 2006; VanPatten, 1996)

(5) creating production activities that 'trap' certain language features to make them essential or useful for meaning (Loschky & Bley-Vroman, 1993) when they could otherwise be avoided or circumnavigated, promoting a (highly-)structured version of 'pushed output'.

Steps 1-3 involved defining a body of language knowledge. Until the very recent policy change (DfE, 2025/2023/2022; see Marsden & Hawkes, 2024 for an account of the change), this body of knowledge was relatively loosely defined in a National Curriculum (DfE, 2013) and a GCSE content (DfE, 2022/2014), both with only high-level descriptions of a very wide range of grammar and an ill-defined lexicon. There was little evidence about how much can be learnt in the time available, reliably across learners with a wide range of cognitive and affective individual differences. We needed to make decisions about: how many words and how much grammar could be taught per week, how many times this language could be revisited (explicitly and incidentally) across the five-year course. To the best of our knowledge, no published material had done this at a fine level of detail for French, German, or Spanish (Dudley et al., 2024; Marsden & Hawkes, 2023).

Steps 4-5 involved creating resources that foregrounded practice, including comprehension and production of features essential to understand or express meaning, in different modes (comprehension/production), modalities (oral/written), with a communicative or thematic need or domain of use, in different linguistic distributions (familiar grammar with new words; new grammar with familiar words). We used MultlingProfiler to make informed decisions about when to use words that were 'on' or 'off' our word lists, to support consolidation or inferencing respectively.

Other kinds of activities, such as topic-themed roleplays or structured and free written production were already widely and confidently used, and so less of a priority to create given our constraints. By the end of the first three years, 94.2% of the teachers in the network reported that the resources supported their teaching (Marsden et al., 2023, p.39).

During the pandemic, in addition to these resources, Rachel led and co-ordinated the work of many of the specialist teachers video-recording hundreds of these lessons for

national free resources (Oak National Academy, https://www.thenational.academy/#teachers) for children unable to attend school.

2.4 Phase Four (years 3-4.25): Working within and beyond the network

Year three's activities included bespoke PD for Lead and Hub schools on issues that the teachers themselves identified as challenging or needing support. After three years, the Centre had delivered PD with 290 teachers from the initial network of 45 schools. Reach in year three was also extended beyond that network, via two main routes: Lead Schools opened their full day conferences to more schools averaging 15 per Hub, and several thousand teachers beyond the network attended stand-alone talks, webinars, conferences, and workshops delivered by NCELP or school-based colleagues.

For the fourth year, we sought permission from the DfE to work with teachers beyond the initial network. To achieve this, 12 teachers from Lead Schools delivered an online course of five 2.5-hour sessions. Delivery formats were a synchronous online course; an asynchronous online course; and a blended course with synchronous and asynchronous components. The course covered six themes (curriculum design, phonics, vocabulary, grammar, culture and cultural capital, assessment) and was delivered by the 12 teachers to 859 other teachers. However, it became evident that committing to 12.5 hours was not always possible for teachers. We therefore changed operations to offer PD via talks and workshops for clusters of schools[7], ranging from an hour to full days and multi-part twilight (after school hours) programmes[8]. Although, as noted above, longer PD is desirable, shorter, narrow-focused PD has been observed to be equally effective when delivering messages about specific research-informed principles (Cordingley et al., 2015; Kennedy, 2016). These wider dissemination activities involved approximately 1,825 teachers. Combined with the five-session courses, 2,684 teachers took part in PD in the final 1.25 years. Additionally, the 12 specialist teachers held open-door events, so others could observe lessons and participate in question-and-answer opportunities.

PART 3 CHALLENGES, LEGACIES, AND LESSONS

---

[7] Currently, in England schools collaborate in different ways in groups such as Local Authorities, Multi Academy Trusts, Teaching School Hubs, and Teaching Alliances.
[8] We sought agreement from the DfE that these would all contribute to the DfE's KPIs, for which non-compliance could incur financial penalties.

Work at research-policy-practice interfaces is not straightforward. There were the predictable challenges of working with schools at a time of sector-wide difficulties, some related to the pandemic but others more systemic such as teacher recruitment and retention. To illustrate, we were notified of 103 staff changes in the 45 schools between 2019-2021 (Marsden et al., 2023: 12). Challenges also arose from working—on a project without a blueprint—with many colleagues on very short contracts. For example, we only slowly realised the scale and complexities of resource development as it progressed. Other experiences can perhaps provide more insight for generalisable 'lessons', and we give a frank account of some of these.

3.1 Tensions between research-informed support and teacher agency

On the one hand, our PD and resources were welcomed by many, as we responded to practitioners' requests, engaging in co-production of research-related artefacts. We hoped that resources would stimulate new discourse, and we actively welcomed—and continue to welcome—alternatives, or different versions with 'more of x', 'less of y'. Resources and PD materials were released in editable formats, so that teachers and teacher educators could tailor them. Crucially, they could also ignore them, and of course some did, quietly making local decisions. Data—including feedback directly from teachers, from the DfE-required surveys (Marsden et al., 2023), and from our surveys about teacher learning and attitudes to research (Marsden & Dudley, in progress)—showed that many teachers reported that the PD and resources broadened thinking. This is perhaps neatly captured by one teacher asking: "why has no-one told us about this research [about high frequency vocabulary] before now?" Perhaps more importantly, many *outside* the network voluntarily adapted ideas to their own practice. OASIS summaries apparently stimulated ideas and autonomous behaviour, too: Alferink and Marsden (2023) found that 78% of 220 teachers agreed that OASIS summaries "helped them develop new ideas" and 88% indicated that they "intended to keep using OASIS summaries after the PD course".

On the other hand, a small but vocal number of stakeholders (mainly, some teacher educators, commercial materials writers, one or two prolific and highly influential commercial PD providers, and some academics with specialisms in literature/culture) publicly challenged the idea of a language-driven curriculum and resources to develop components of language proficiency. We suggest four possible reasons for some of this negative commentary.

First, once research ideas are operationalised in materials, exemplars become tangible for critique; a healthy outcome and arguably something that all researchers claiming to work at research-practice interfaces should deliver and welcome.

Second, our work was *government*-funded. Some elements of the PD (such as the value of providing models, scheduled practice, controlling the amount and type of information, some direct instruction, and whole class questioning to check learning) broadly aligned with top-down policy of the time that foregrounded ideas from mainstream educationalists and cognitive psychologists (e.g., Hattie, 2009; Rosenshine, 2012; Sweller, 1994). In fact, these sources were not cited in our PD or resources[9], as we drew on *applied linguistics* research and teacher expertise. Nevertheless, some commentators, mainly on social media, webinars, or in a teacher-oriented magazine, swept our work up with critique of the Conservative government's education policy (Riordan & Smith, 2022). Similarly, given the positive data summarised above—from teachers within and beyond the network—we were disappointed to see our work described as "stream-lining and unifying… reducing the diversity of methods, materials and approaches… with all the potential for limiting teacher autonomy" (Lanvers & Graham, 2022[10], p. 231). Arguably, this critique—ironically a researcher-perspective, articulated seemingly 'on behalf of' teachers —itself reflects a desire to stifle certain types of innovation, thus potentially narrowing routes to experimentation and development. These and others' perceptions of 'homogenising' were perhaps perpetuated by the fact that NCELP was government-funded at high value, even though most of money went to the Lead Schools themselves. Such perceptions may also result from a feeling of exclusion and/or from the sheer pace of a government-funded project that didn't give sufficient time for some commentators to explore or understand all aspects of the work as it evolved.

Third, our selection of research introduced 'biases', as intermediaries are necessarily selective, reflecting the very concerns and caveats laid out by the founders of OASIS (Marsden & Kasprowicz, 2017). Our selection was driven by criteria as explained above and our scope was constrained by time and funds; but no selection of research can ever please all. For example, as noted above, we did not prioritise research that relies *heavily* upon either implicit learning mechanisms (input floods or CLIL) or every individual's analytic ability to induce their own reliable and meaningful pattern learning. However, such learning

---

[9] Other than those mentioned above, in which we were asked to situate our work in a familiar frame of reference for teachers.

[10] For transparency, both these authors were part of NCELP, the first author for a few months, the second for two years.

mechanisms and pedagogical approaches were certainly not simply dismissed as entirely irrelevant. We acknowledged other accounts of learning, and often referred to the fact that we were foregrounding 'evidence of likely *trends* that are likely to be relevant given *this kind of teaching context*, for *most* learners, *most* of the time'. 'Biases' were perhaps simply the consequence of having to prioritise given the constraints of: our remit; the limited time available to learners; the characteristics of this very broad population of learners; and given that some research-informed innovations, such as strategy-instruction, had already had opportunities to inform practice.

A fourth concern was that our work disrupted funding structures in school language education: Our materials and PD were free, thus threatening income for independent commercial providers of PD and materials. At least one major publisher of school materials lobbied government to stop funding resource development. Also, a vocal commercial PD provider frequently posted unevidenced critiques of very short extracts or isolated ideas from our resources on social media, reaching thousands of followers.

We admit that unevidenced critique was often upsetting and ultimately divisive, given our shared ambition of improving learners' and teachers' experiences. The critiques—each with *different* motivations—exerted additional pressures to sustain our introspective check on our integrity. Working under a large government contract and with policymakers meant that opportunities to compromise or uphold our integrity presented themselves more than at other times in our careers. This situation is perhaps best illustrated when the opportunity to apply for a subsequent contract—thankfully permitting a wider remit, such as community and home languages—presented itself, potentially amounting to £19 million. Although ultimately unsuccessful, in our bid we chose to not compromise our ethical values or research-informed principles. This involved, for example, resisting covert and overt pushes to collaborate with specific national/international government-backed cultural institutions or specific, politically favoured school trusts.

Generally, we now feel that increased and enriched explicit discussion about the *language* content of curricula and pedagogy has been a positive outcome in the longer term. We still witness new discourse filtering through networks; and in several cases, what started as criticisms from some (about, for example, phonics and high frequency vocabulary) has steadily morphed into these ideas being widely incorporated into commercial materials and mainstream PD.

3.2 Need for data and evaluations.

In terms of evaluating teachers' engagement and attitudes throughout the programme, the DfE asked us to collect basic survey data on teachers' satisfaction to evidence the funder's KPIs (Marsden et al., 2023). For example, over the 4.25 years, 97% of teachers who attended PD reported an increased understanding of research-informed pedagogy, curriculum design, and assessment. These data are crude and probably affected by self-selection and potentially also self-confirmation bias given the financially incentivising contracts formally underpinning the teachers' involvement. Independently, however, we gathered more nuanced data from attendees on the five-session course in year four about: 1) engagement with research; 2) attitudes to research; 3) opinions about OASIS summaries; 4) knowledge about research covered by the PD; 5) engagement with the PD course. 1058 teachers provided data, of which 404 provided pre-post data (Marsden & Dudley, in progress). However, as Mujis et al. (2014: 248) observed, "for teachers, one of the most powerful sources of feedback comes from how students respond to the changes they make to their practice" and many teachers told us that their students and they themselves were benefiting from the PD and resources. These qualitative comments were plentiful, often provided voluntarily via WhatsApp, texts, emails, conversation, Facebook, or in the optional open-text field in end-of-course feedback forms. Many of these quotations from teachers are publicly available here. Perhaps the most powerful testimony, though, is that hundreds of teachers (possibly thousands, given the 700,000+ downloads of resources to date and the numbers in two independent Facebook groups) *beyond* the initial network of teachers have adopted and/or adapted NCELP curricula and resources.

In terms of the evidence underlying our PD and resources, we drew mainly on relatively small-scale studies investigating components (e.g., vocabulary or grammar teaching) or meta-analyses thereof. But we need holistic evaluations of the effects of our PD and/or resources on learners' cognitive and/or affective outcomes, from which generalisable recommendations could be inferred (Norris, 2016). The DfE did not fund an evaluation[11], and yet, as Marsden et al. (2022) observed in their final report, "it might help speed up full engagement with a largely new pedagogy if research evidence from a large-scale randomised control trial of the holistic pedagogical approach (and associated assessments of learning) were available prior to asking schools to change curriculum, pedagogy, and assessments." (p. 11). However, evaluations of holistic interventions are notoriously difficult to undertake reliably and validly, in such a way that informs both practice and *theory* (Marsden, 2006;

---

[11] Nor have the ESRC or Nuffield to date.

Marsden & Torgerson, 2012). Such evaluations are additionally difficult when the intervention comprises multiple components, and so choosing a fair comparison is almost impossible and overlapping components make identifying causal levers challenging (though see Minett, in progress; Graham et al., 2017; Woore et al., 2018).

Another approach is to manipulate one or a very small number of *components* in (semi-)simulated interventions—experimental and comparison group(s)—both realistic for the context, and teachers can be assigned to the condition that most aligns with their normal behaviours (Saito & Loewen, 2018). For example, our experience told us that we need studies in our context on: massed versus spaced phonics practice; deductive versus inductive grammar practice; long-term impact of texts with different lexical coverages (more versus less familiar vocabulary). Indeed, single institution experiments—which offer potential for extending to larger trials—are recommended by educationalists (e.g., Mujis, 2014: 250-251) as they can both support teacher development and inform understanding of educational effectiveness. Such small-scale studies constitute the core of Instructed Second Language Acquisition research, but the populations—usually learners of English and/or students in Higher Education—need diversifying to other contexts (c.f., Kasprowicz & Marsden, 2017).

To this end, during our project's second-third years, we encouraged specialist teachers to undertake small-scale action research. A handful of teachers met with researchers (Kasprowicz, Marsden, Woore) to discuss their interest in doing this on topics such as phonics or grammar teaching. We set aside funding, offering mentorship and assistance (e.g., data collection or analysis). However, the teachers did not find time to pursue these projects, partly because Covid19 brought unforeseen pressures but also because the teachers expressed concerns about the time investment. Our experience did not, therefore, clearly reflect recommendations in the literature that action research can be an avenue for PD, at least not as an additional voluntary undertaking.

We jointly learned that innovation, even when research-informed, quite rightly elicited demands for 'proof of effectiveness' from a few commentators (e.g., an influential commercial PD provider; a petition from the All-Party Parliamentary Group campaigning about revisions to the GCSE). On the other hand, we observed that a (or several) *status quo* has perpetuated without such demands for proof, even in plain sight of negative evidence (e.g., falling numbers of students; teacher dissatisfaction). This asymmetrical level of demand for 'proof of effectiveness' perhaps simply reflects general resistance to change (observed, by Minett, in progress, among primary teachers). But it is sobering to note that many, or indeed most, approaches, such as topic-driven slot-and-fill activities—sometimes called 'sentence-

builders'—have perpetuated extensively in this context for decades *without* rigorous outcome-focused trials.

In sum, we hope that both holistic and componential evaluations of language-driven pedagogies will be funded, so that teacher and student outcomes can be fairly and usefully compared to those of other approaches.

3.3 Sustaining and consolidating innovation

Effective research-policy-practice interfaces inevitably involve "iterative processes as teachers learn new knowledge and refine existing skills, try things out in practice, work out what is working and not working for students, revisit conceptions and misconceptions, and try again" (Mujis et al., 2014: 248). As such, we briefly consider what has been sustained post government funding, illustrating four of the five factors that Florian (2000: 4) proposed as contributing to sustained change.

a) "*methods or practices that teachers experienced as effective in accomplishing school goals*": The resources have been used and adapted by individual teachers, schools, and clusters of schools. Two Facebook pages were independently created in 2022 by teachers using NCELP resources, involving approximately 2,000 teachers. The MultilingProfiler https://multilingprofiler.net/ is being used to lexically profile thousands of texts for teaching. Furthermore, some of the largest and/or most successful chains of schools in England (e.g., *United Learning*; *Twyford CofE Academies Trust*; *Westcountry Schools*) are now using, adapting, and creating their own versions of language-driven curricula. Other evidence that teachers experienced effective methods or practices is that commercial entities chose to use NCELP-inspired materials. For example, a language education technology company, Languagenut, used our word lists and vocabulary practice to create widely used activities. Although we insisted on free availability for those, perpetual free availability was not possible for all materials: In summer 2023, Languagenut bought an exclusive licence for ten years to use our online app providing input-based gamified practice (Gaming Grammar, funded by DfE and other funds). Another tech company, Sanako UK, worked with us in 2022-23 to develop some online activities. These examples illustrate what can happen when public funding ceases: Whilst open scholarship is the ideal, maintenance costs mean that commercial paywalls can become necessary (see discussion by Marsden & Morgan-Short, 2023). On a brighter note, one of these companies (Sanako) now sponsors the national

Spelling and Translation Bee competitions, which built on NCELP's work on vocabulary selection.

b) "*school principals who effectively promoted, supported and managed change*". School leaders, both within and beyond our network, supported and sustained change such that teacher turnover during our project did not erode participation or engagement. Also, NCELP teachers were supported in bringing ideas to their new schools and school clusters.

c) "*political support for new practices from district and, if possible, state levels*". After our funding ceased, two major government investments aimed to sustain some of the work. In January 2024, Hawkes was awarded a large tender to work with teachers to create materials for the Oak National Academy, https://www.thenational.academy/#teachers, using NCELP's curriculum and revised resources. Second, the DfE awarded a very large contract to 'UCL Consultants'[12]. This was intended to build on and extend NCELP's work including maintaining NCELP's resources. Out of the 15 Lead Schools under this new arrangement, at least four use NCELP's schemes of work and resources and incorporate them into the PD they deliver to their five-seven partner Hub schools. Finally, several elements have so far survived the recent change in the UK from Conservative to Labour government in July 2024, including: the two investments just mentioned; TSC (2016); and changes to the national GCSE examinations for 16-year-olds (DfE 2025/2023/2022; Finlayson et al., 2024; Marsden et al., 2023; Marsden & Hawkes, 2024). Finally, though not strictly-speaking 'political' support, the national BBC Bitesize and Teacher Education teams created two free audiovisual teacher education resources with NCELP teachers demonstrating and describing approaches to phonics and grammar teaching.

d) "*continual, high-quality professional development and/or assistance for staff*". In addition to the investments described above, in April 2023, our revised unit ('LDP') has provided PD (though necessarily now at a small fee) delivered by four teachers to approximately 200 teachers. Many teachers have attended talks and workshops delivered by Hawkes. Events

---

[12] The new contract, which funded a 'National Consortium for Language Education' (UCL) and two established cultural institutions (the Goethe Institute and the British Council), had a broader remit. Initially contracted to establish at least 25 Hubs, after about one year, the DfE cut funding to 15 Hubs. Then, from April 2025 (18 months after PD started), funding for the Hubs was cut entirely. These cuts are perhaps indicative of the challenges of establishing and nurturing genuine collaborations that *truly* and *simultaneously* serve the interests of *all three* stakeholder-types in a policy-research-practice alliance.

planned for 2024-25 include two large school chains (*Swift* and *Harris Federation*) who requested support for their own language-driven pedagogies.

3.4 Personal lessons from our partnership

What happens to a close researcher-teacher partnership once major funding stops? We continue to collaborate on the PD unit 'LDP', established thanks to university funds but which must become financially self-sustaining (a loss to open scholarship). We also collaborate through Oak National Academy as Marsden sits on the advisory group.

Generally, however, we have returned to our university- and school-based 'homes', each with different infrastructures and incentivisation, though as changed professionals. Marsden perceives effects of the partnership in various dimensions of her work: research questions, designs, sampling decisions, collaborators, and interpretations of research. For example, the absence of evidence we found about knowledge and proficiency after five years of instruction in this context prompted a large study to address this gap (Dudley & Marsden, under contract). Hawkes carries the work forward more concretely, leading the production of resources for ages 7-16. Her advisory and PD work continues to be influenced by the research she became familiar with during NCELP and, perhaps more importantly, she sustains the openness to research that she had at the start of our partnership.

In addition to these positives, we share a (bleaker?) realism. We experienced how political strategizing—at national and even international levels—and reactionary pressure groups (e.g., some publishers, awarding organisations, or parliamentary groups) can exert influence over innovation and funding decisions. Our reflections in this piece have made it clear that the government funding, in our *three-way* collaboration between policymakers-practitioners-researchers, certainly added complexities and challenges. But, on the other hand, the central funding gave affordances that would not have been available with an entirely grassroots or researcher-led initiative. These wide-ranging affordances include several at a personal-professional level. Emma has learnt more about the extent to which research can be of genuine and lasting relevance and use to policymakers and practitioners and, like Laufer (2023), has a greater awareness of the constraints facing policymakers and practitioners, both when they interpret research, are involved in it, or wish to operationalise/integrate/accommodate aspects of it into policy/practice. Rachel acquired greater criticality about what research can offer and greater understanding that many challenges faced by practitioners constitute enormous research agendas.

Another lesson learned is that we both continue to witness 'over-assimilation' (Hammerness et al. 2005, cited by Mujis et al., 2014; p. 250), where new ideas are only interpreted *within* existing ideas and so only superficial or no deep changes are made to thinking or practice. For example, in webinars and others' PD we have heard use of terms such as 'revisiting', 'high frequency vocabulary', 'processing instruction', 'knowledge', or 'skill', without appreciation of concrete consequences for practice (e.g., the need to sequence a curriculum to operationalise these ideas). We also still see the simplistic dichotomisation of harmfully pitting a focus on language 'against' cultural enrichment and communicative interaction. These observations are, we think, a consequence of the lack of continuity of "*intensive* and *challenging* professional learning experiences to engage teachers in activities and dialogue to […] identify the difference between the beliefs they hold and the beliefs underpinning the new ideas" (Mujis et al., 2014; p. 247-248, italics ours).

On the whole, however, we believe that our close partnership was fruitful for us—and we hope others—to better understand some key ingredients of work at research-policy-practice interfaces: striving for a shared conceptualisation of the challenges facing policymakers, researchers, and teachers; iterative and reactive approaches to identifying and sharing new concepts and findings from research; PD in which researchers and teachers had opportunity to reflect on these and their contextual relevance; and the production, piloting, and refinement of editable resources to contextualise these ideas and facilitate autonomy for teachers in owning them. A reviewer asked us: "Would you do it again?". It's hard to say, but we think overall, yes, we would. But with some lessons learned, it would of course be different.

References

Alferink, I. & Marsden, E. (2023) *OASIS: one resource to widen the reach of research in language studies. Innovation in Language Learning and Teaching*. pp. 946-952. ISSN 1750-1229

Andringa, S., Olsthoorn, N., van Beuningen, C., Schoonen, R., & Hulstijn, J. (2012). Determinants of success in native and non-native listening comprehension: An individual differences approach. *Language Learning*, *62*, 49–78. https://doi.org/10.1111/j.1467-9922.2012.00706.x

Churchward, D. (2019). *Recent trends in modern foreign language exam entries in anglophone countries*. Ofqual. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/844128/Recent_trends_in_modern_foreign_language_exam_entries_in_anglophone_cou

ntries_-_FINAL65573.pdf

Coffey, S. (2016). Choosing to study modern foreign languages: Discourses of value as forms of cultural capital. *Applied Linguistics*, *39*(4), 462–480. https://doi.org/10.1093/applin/amw019

Colleen, I., & Duff, J. (2024) Language Trends England. *British Council*

Consoli, S., & Dikilitaş, K. (2021). Research Engagement in Language Education. *Educational Action Research*, 29(3), 347–357. https://doi.org/10.1080/09650792.2021.1933860

Cordingley, P., Higgins, S., Greany, T., Buckler, N., Coles-Jordan, D., Crisp, B., Saunders, L., & Coe, R. (2015). *Developing great teaching: Lessons from the international reviews into effective professional development*. Teacher Development Trust. https://tdtrust.org/wp-content/uploads/2015/10/DGT-Full-report.pdf

Curcin, M., & Black, B. (2019). *Investigating standards in GCSE French, German and Spanish through the lens of the CEFR*. Ofqual. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/844034/Investigating_standards_in_GCSE_French__German_and_Spanish_through_the_lens_of_the_CEFR.pdf

Department for Education, 2013. National curriculum in England: Languages programme of study. https://www.gov.uk/government/publications/national-curriculum-in-england-languages-progammes-of-study/national-curriculum-in-england-languages-progammes-of-study

Department for Education, 2019. *Early career Framework*. London: HMSO. January 2019.

Department for Education. (2022, December 2). *GCSE modern foreign languages*. https://www.gov.uk/government/publications/gcse-modern-foreign-languages. (First published 2014, April 9).

Department for Education. (2025, February 24). *GCSE French, German and Spanish subject content*. https://www.gov.uk/government/publications/gcse-french-german-and-spanish-subject-content. (First published 2022, January 14; revised February 20).

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181–199. https://doi.org/10.3102/0013189X08331140

Dudley, A., & Marsden, E. (2024). The lexical content of high-stakes national exams in French, German, and Spanish in England. *Foreign Language Annals*, *57*, 311-338. https://doi.org/10.1111/flan.12751

Dudley, A., & Marsden, E. (under contract). *Components and correlates of language proficiency: The case of adolescent classroom learners of French and Spanish in England*. Multilingual Matters.

Dudley, A., Marsden, E., & Bovolenta, G. (2024). A Context-Aligned Two Thousand Test: Toward estimating high-frequency French vocabulary knowledge for beginner-to-low intermediate proficiency adolescent learners in England. *Language Testing, 41*(4), 759-791. https://doi.org/10.1177/02655322241261415

Education Endowment Foundation. (2021). *What are the characteristics of effective teacher professional development? A systematic review and meta-analysis*. https://d2tic4wvo1iusb.cloudfront.net/documents/pages/Teacher-professional-development.pdf?v=1633521331

Finlayson, N., & Marsden, E. (2022). Assessing depth of word knowledge of beginner learners of French, German, and Spanish aged 11-14 in England. In S. W. Chong, & H. Reinders (Eds.), *Learner-Centred Approaches to Language Assessment* Palgrave Macmillan.

Finlayson, N., Marsden, E., & Anthony, L. (2023). Introducing MultilingProfiler: An adaptable tool for analysing the vocabulary in French, German, and Spanish texts. *System*, *118*, Article 103122. https://doi.org/10.1016/j.system.2023.103122

Finlayson, N., Marsden, E., & Hawkes, R. (2024). Creating and evaluating corpus-informed vocabulary lists for adolescent, beginner-to-low-intermediate learners of French, German, and Spanish, *Language Teaching Research. https://doi.org/10.1177/13621688241288877*

Florian, J. (2000). Sustaining Education Reform: Influential Factors. Office of Educational Reform and Improvement U.S. Department of Education. https://files.eric.ed.gov/fulltext/ED453583.pdf

Garon–Carrier, G., Boivin, M., Guay, F., Kovas, Y., Dionne, G., Lemelin, J.-P., Séguin, J. R., Vitaro, F., & Tremblay, R. E. (2016). Intrinsic motivation and achievement in mathematics in elementary school: A longitudinal investigation of their association. *Child Development, 87,* 165–175. https://doi.org/10.1111/cdev.12458

Graham, S. (2024) Engaging with and in research for classroom practice: Supporting the beginning languages teacher. Presentation to the *Annual Conference of the Initial Teacher Education and Training* Special Interest Group of the UK's Association for Language Learning, 16th July 2024.

Graham, S., Courtney, L., Marinis, T., & Tonkyn, A. (2017). Early language learning: the impact of teaching and teacher factors. *Language Learning, 67*: 922-958. https://doi.org/10.1111/lang.12251

Hattie, J. (2009). *Visible learning. A synthesis of over 8000 meta-analyses relating to achievement.* London: Routledge.

Hawkes, R., & Lillington, C. (2016). *Viva! AQA GCSE (9-1) Spanish foundation student book*. Pearson Education.

Jeon, E. H., & Y. In'nami, Y. (Eds.) (2022), *Understanding L2 proficiency: Theoretical and meta-analytic investigators*. John Benjamins Publishing Company.

Joint Council for Qualifications (2024). GCSE results Summer 2024
https://www.jcq.org.uk/gcse-level-1-and-level-2-results-summer-2024/?mc_cid=ae080cfe23

Kasprowicz, R. E., & Marsden, E. (2018). Towards ecological validity in research into input-based practice: Form spotting can be as beneficial as form-meaning practice. *Applied Linguistics, 39*(6), 886-911. https://doi.org/10.1093/applin/amw051

Kasprowicz, R.E., Marsden, E., & Sephton, N. (2019). Investigating distribution of practice effects for the learning of foreign language verb morphology in the young learner Classroom. *The Modern Language Journal, 103*, 580-606. https://doi.org/10.1111/modl.12586

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research*, *86*(4), 945–980. https://doi.org/10.3102/0034654315626800

Lanvers, U., & Graham, S. (2022). Can we design language education policy and curricula for a motivated learner? Self-determination theory and the UK language crisis. *The Language Learning Journal*, *50*(2), 223–237. https://doi.org/10.1080/09571736.2022.2046353

Laufer, B. (2023) From research to a national curriculum: The case of a lexical syllabus. In G. Erickson, C. Bardel, and D. Little (Eds.) *Collaborative Research in Language Education: Reciprocal Benefits and Challenges*. Berlin, Boston: De Gruyter Mouton, pp. 151-164. https://doi.org/10.1515/9783110787719-011

Lightbown, P. (1985). Great Expectations: Second-language acquisition research and classroom teaching, *Applied Linguistics*, 6, 2, 173-189, https://doi.org/10.1093/applin/6.2.173

Loschky, L. & Bley-Vroman, R. (1993) Grammar and task-based methodology. In G. Crookes and S. Gass (Eds.) *Tasks and language learning: Integrating theory and practice.* Clevedon, England: Multilingual Matters, pp. 123–167.

Lynch, S. J., Pyke, C., & Grafton, B. H. (2012). A retrospective view of a study of middle school science curriculum materials: Implementation, scale-up, and sustainability in a changing policy environment. *Journal of Research in Science Teaching, 49(3),* 305–332. https://doi.org/10.1002/tea.21000

Marsden, E. (2006). Exploring input processing in the classroom: An experimental comparison of processing instruction and enriched input. *Language Learning*, *56*(3), 507-566. https://doi.org/10.1111/j.1467-9922.2006.00375.x

Marsden, E. (2007), Can educational experiments both test a theory and inform practice? *British Educational Research Journal*, 33: 565-588. https://doi.org/10.1080/01411920701434094

Marsden, E. & Dudley, A. (in progress). Teachers' attitudes to and engagement with research and the roles of a professional development course using OASIS summaries.

Marsden, E., & Hawkes, R. (2023). Situating practice in curriculum design in school foreign language education. In Y. Suzuki (Ed.) *Practice and Automatization in Second Language Research: Perspectives from Skill Acquisition Theory and Cognitive Psychology.* Chapter 4

Routledge. pp 89-118

Marsden, E. & Hawkes, R. (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

Marsden, E., Hawkes, R., Earnshaw, L., Hobson, V., & NCELP team (2023) NCELP Final Report on the National Centre for Excellence for Language Pedagogy Accessed https://eprints.whiterose.ac.uk/200637/1/NCELP_Final_Report_9_April_2023.pdf

Marsden, E., & Morgan-Short, K. (2023), (Why) are open research practices the future for the study of language learning? *Language Learning, 73*, 344-387. https://doi.org/10.1111/lang.12568

Marsden, E., & Torgerson, C. J. (2012). Single group, pre- and post-test research designs: Some methodological concerns. *Oxford Review of Education*, *38*(5), 583–616. https://doi.org/10.1080/03054985.2012.731208

Marsden, E., Dudley, A., & Hawkes, R. (2023). Use of word lists in a high-stakes, low-exposure context: Topic-driven or frequency-informed. *The Modern Language Journal*, *107*(3), 669–692. https://doi.org/10.1111/modl.12866

Marsden, E. & Kasprowicz, R. (2017). Foreign language educators' exposure to research: Reported behaviours, exposure via citations, and a call for action. *The Modern Language Journal. 101* (4), 613–642

Minett, E. (in progress). Comme ci, comme ça? A comparison of two approaches to teaching French to 7-8 year olds in England. Ongoing PhD thesis. University of York.

Mitchell, R. (2002). Foreign language education in an age of global English. Centre for Language in Education Occasional Papers 57. University of Southampton.

Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014) State of the art – teacher effectiveness and professional learning, *School Effectiveness and School Improvement, 25*:2, 231-256, DOI: 10.1080/09243453.2014.885451

Myles, F., Hooper, J. & Mitchell, R. (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning, 48,* 323–363.

National Association of Language Advisers (2020) The languages curriculum and disadvantaged students. Survey report October 2020. https://nala.org.uk/wp-content/uploads/2023/11/NALA-Report-on-socio-economic-deprivation-and-MFL-2020-Full-report.pdf

Norris, J.M. (2016), Language program evaluation. *The Modern Language Journal*, *100*, 169-189. https://doi.org/10.1111/modl.12307

Ofqual (2019). Inter-subject comparability in GCSEs.
https://www.gov.uk/government/news/inter-subject-comparability-in-gcse-modern-foreign-languages

Ofsted (2008). *The changing landscape of languages – an evaluation of language learning 2004/07*. Report no. 070053. https://dera.ioe.ac.uk/id/eprint/8192/.

Ofsted (2011). *Modern Languages: achievement and challenge 2007-2010.* Report no. 100042 https://www.gov.uk/government/publications/modern-languages-achievement-and-challenge-2007-to-2010

Ofsted (2021). Research review series: languages.
https://www.gov.uk/government/publications/curriculum-research-review-series-languages/

Ofsted (2023) The School Inspection Handbook
https://www.gov.uk/government/publications/school-inspection-handbook-eif/school-inspection-handbook-for-september-2023

Ortega, L. (2013), SLA for the 21st century: Disciplinary progress, transdisciplinary relevance, and the bi/multilingual turn. *Language Learning, 63*: 1-24. https://doi.org/10.1111/j.1467-9922.2012.00735.x

Piper, B, Simmons Zuilkowski,S., Dubeck , M., Jepkemei , E., King, S., (2018), Identifying the essential ingredients to literacy and numeracy improvement: Teacher professional development and coaching, student textbooks, and structured teachers' guides. *World Development,* 106, 324-336, https://doi.org/10.1016/j.worlddev.2018.01.018

Pringle, R.M., Mesa, J. & Hayes, L. Meeting the demands of science reforms: a comprehensive professional development for practicing middle school teachers. *Res Sci Educ* 50, 709–737 (2020). https://doi.org/10.1007/s11165-018-9708-9

Reeves, J. (2010). Teacher learning by script. *Language Teaching Research*, *14*(3), 241-258. https://doi.org/10.1177/1362168810365252

Riordan, T., & Smith, C. (2022). Who determines teacher agency? Are we being constrained by an ideological agenda to teach in a certain way? *Languages Today*, *Autumn 2022*(42), 30. https://www.all-languages.org.uk/research-practice/languages-today/

Rose, H., & Mckinley, J. (2022). May I see your credentials, please? Displays of pedagogical expertise by language teaching researchers. *The Modern Language Journal, 106*(3), 528-546.

Rosenshine, B. (2012) Principles of instruction: Research-based strategies that all teachers should know. *American Educator, Spring, 12-39.*

Sato, M. (2023) Navigating the research–practice relationship: Professional goals and constraints. *Language Teaching:*1-16. https://doi.org/10.1017/S0261444823000423

Sato, M., & Loewen, S. (2018). Metacognitive instruction enhances the effectiveness of corrective feedback: Variable effects of feedback types and linguistic targets. *Language Learning, 68:* 507-545. https://doi.org/10.1111/lang.12283

Shu, D., Yang, S., & Sato, M. (2024) Marrying collective wisdom: Researcher-practitioner collaboration in developing ELT textbooks, *Applied Linguistics*, 45(1), 41–64, https://doi.org/10.1093/applin/amac074

Sweller, J. (1994) Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction. 4, 4, 295-312*

Taylor, J.A., Getty, S. R., Kowalski, S. M., Wilson, C. D., Carlson, J., & Van Scotter, P. (2015). An efficacy trial of research-based curriculum materials with curriculum-based professional development. *American Educational Research Journal, 52*(5), 984–1017. https://doi.org/10.3102/0002831215585962

Taylor, F. & Marsden, E. (2014), Perceptions, attitudes, and choosing to study foreign languages in England: An experimental intervention. *The Modern Language Journal, 98*, 902-920. https://doi.org/10.1111/modl.12146

Teaching Schools Council (TSC) (2016). Review of MFL Pedagogy. *Teaching Schools Council. https://www.york.ac.uk/education/research/language-driven-pedagogy/resources/modern-foreign-languages-pedagogy-review/*

Timperley, H. (2011). Realizing the power of professional learning. London: Open University Press.

VanPatten, B. (1996). *Input processing and grammar instruction in second language processing.* Ablex.

Watson, P., Hobson, V., Marsden, E., & Hawkes, R. (2021) *Connecting research into pedagogy with foreign language teaching: A perspective from NCELP.* Impact Journal of the Chartered College of Teaching, *13*. ISSN 2514-6955

Wingate, U. (2016). Lots of games and little challenge – a snapshot of modern foreign language teaching in English secondary schools. *The Language Learning Journal*, *46*(4), 442–455. https://doi.org/10.1080/09571736.2016.1161061

Woore, R., Graham, S., Porter, A., Courtney, L., & Savory, C. (2018). Foreign language education: unlocking reading (FLEUR) - A study into the teaching of reading to beginner learners of French in secondary school. https://ora.ox.ac.uk/objects/uuid:4b0cb239-72f0-49e4-8f32-3672625884f0.

Zhang, S., & Zhang, X. (2022). The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis. *Language Teaching Research*, *26*(4), 696–725. https://doi.org/10.1177/1362168820913998