

What are the possibilities and limitations of small-scale, mediated classroom interventions? Two cases from Modern Foreign Languages.

Laura Molway (University of Oxford), Trevor Mutton (University of Oxford), Robert Woore (University of Oxford) and Katherine Porter

Abstract

This chapter considers two case studies exemplifying different models of small-scale, mediated intervention research in Modern Foreign Language (MFL) classrooms. By ‘mediated’, we mean that they are thoroughly grounded in both theoretical and practical perspectives, such that practical concerns are mediated by research understandings, and vice versa. Both studies focus on the development of assessment practices at a time when recent changes to the National Curriculum in England have resulted in the removal of the previous well-established assessment framework. The first was carried out by a teacher as part of an examined Master’s-level research project, the second by university academics working collaboratively with other practitioners within a multi-level Initial Teacher Education partnership. Both studies illustrate how a thoroughly research-informed approach can be useful in addressing practical classroom problems. We also argue that such mediated, small-scale studies can be a valuable first step in a wider process of systematically developing and evaluating pedagogical innovations – provided that sufficient opportunities for dissemination are found, which may currently be limited in the case of practitioner research. Finally, we highlight the value of these studies as vehicles for professional learning, as one way of developing teachers’ research literacy.

Key words

Teacher practitioner research; Modern Foreign Languages; MFL; classroom-based interventions; assessment framework.

Introduction

This chapter sets out to explore the possibilities and limitations of small-scale, mediated intervention studies taking place in Modern Foreign Language (MFL) classrooms in the UK. We begin by defining and elaborating on the key terms in this statement of aim.

First, we use the term ‘intervention’ to denote a conscious change in classroom practice aimed at improving learning and teaching. For example, a languages teacher might alter their usual practice to address a perceived problem such as low attainment in a particular skill area, such as reading. In a research context, the term assumes that there is some systematic assessment of outcomes before and after the change in practice, in order to provide an answer to the question ‘has the intervention made a difference?’

Second, in relation to our focus on UK MFL classrooms, we would note that teachers in this context are, in theory, able to draw on a strong tradition of classroom-based interventions in the field of Second Language Acquisition (SLA) to inform their practice: see, for example reviews of intervention studies covering a range of different areas by Saito (2012), Lyster, Saito & Sato (2013), Schmitt, (2008) and Boers & Lindstromberg (2012). In many published SLA studies, however, the participants are high proficiency and/or adult learners of English. These participants differ from MFL learners in terms of their age, linguistic proficiency level and motivational profile – partly because the languages learned in classrooms in the United Kingdom (UK) may not have the instrumental or cultural value that English often carries for foreign language learners in other contexts. There have been some published intervention studies conducted in MFL contexts, but so far these have tended to focus on language learning strategy instruction (reflecting the particular interests of the small number of researchers who are active in this context – see for example Macaro & Erler, 2008; Graham & Macaro, 2008). There is much ground left to cover.

Third, our chapter focuses on classroom interventions which are small in scale. Using the terminology of the Education Endowment Foundation (EEF)¹, a major funder of recent classroom-based research, existing interventions range from large-scale ‘effectiveness trials’ (involving 40 or more schools), through medium-scale ‘efficacy trials’ (10 or more schools) to small-scale ‘pilot studies’ aimed at providing proof of concept (typically three or more schools). Finally, there are also practitioner-led studies based in individual classrooms.

¹ <http://educationendowmentfoundation.org.uk>

Many of the latter are conducted each year within the framework of Initial Teacher Education (ITE) and Master's assignment projects. Despite robust criticism (see for example Oancea & Pring, 2008; Thomas, 2016) it is sometimes assumed that large-scale Randomised Control Trials (RCTs) – large-scale because of the need for adequate statistical power – are the 'gold standard' of evidence concerning what works in the classroom. In this chapter we argue that smaller-scale studies which produce rich yet more tentative findings can also be very valuable.

Associated with the issue of scale is the question of the status of the knowledge arising from different kinds of intervention studies, and how the findings are disseminated. Large-scale intervention studies conducted by professional academics will routinely be published. Increasingly, smaller-scale interventions (e.g. Master's or doctoral theses) may also be included in systematic reviews or meta-analyses (in the field of SLA research, for example, see Plonsky, 2011; Hassan, Macaro, Mason, Nye, Smith & Vanderplank, 2005). However, we would question who the readers of these publications might be, beyond the research community itself, and the extent to which such publications are accessible to teachers (even given recent increases in open access to research publications). By contrast, practitioner research often fails to be published or more widely disseminated, and in the UK at least, MFL teachers do not currently have a systematic outlet for this kind of research in the way that teachers of some other disciplines do: for example, the Historical Association journal 'Teaching History' regularly publishes single practitioner case-study research.

Finally, we explain what we mean by a 'mediated' classroom intervention. We argue that the strongest classroom interventions at any scale should be both research-informed and practitioner-informed. To take two 'straw man' examples, we could envisage on the one hand a teacher trying out an intervention in their classroom but failing to draw on existing theories or research; and on the other a professional researcher coming up with an intervention based on theoretical concerns but which is not adequately grounded in practice. In the former case, the teacher may unwittingly be 're-inventing the wheel' or failing to capitalize on what has gone before; in the latter, the intervention may founder simply because it fails to take adequate account of the practical realities of the classroom. By contrast, we can envision a third kind of intervention, grounded in both theoretical and practical perspectives, such that practical concerns are mediated by research understandings, and vice versa. As we shall see, however, this balance is not easy to achieve in practice.

Underpinning this concept of mediation in intervention research is a wider vision of teacher professionalism. We adopt Winch, Oancea and Orchard's (2015) conceptualization of professionalism, which includes an emphasis on research literacy (see also BERA, 2014). That is to say, teachers should be 'equipped to engage with and be discerning consumers of research' and 'equipped to conduct their own research, individually and collectively, to investigate the impact of particular interventions or to explore the positive and negative effects of educational practice' (BERA, 2014, p.11). We would not necessarily expect research findings – even those of large-scale, appropriately mediated intervention studies – to be directly applicable to other classrooms 'off the peg'. Rather, teachers would need to use a critical, informed understanding of a given study's findings and, further, to consider these findings in the light of their contextual knowledge of their own classroom.

The current chapter focuses on two small-scale mediated classroom intervention studies. The factor linking the studies is that each focuses on the development of classroom-based assessment practices in light of recent changes to the National Curriculum in England (DfE, 2013), resulting in the removal of the previous, well-established assessment framework (the National Curriculum 'levels') and the consequent requirement for schools to determine their own assessment frameworks. In discussing these two studies, we will consider questions around the nature of research expertise, professional autonomy, the status of the knowledge produced (and the potential for wider impact) and the potential benefits of teachers' engagement with and in research for their own professional learning.

Background: assessment as a focus for classroom-intervention

Assessment serves a fundamental role in any teaching and learning process as a means to diagnose a pupil's particular strengths and weaknesses, leading to the identification of areas for improvement. Research into the types of assessment that have the potential to promote learning has led to the foregrounding of formative assessment, widely referred to as Assessment for Learning (AfL). Assessment can be considered to be formative when classroom evidence is used to 'adapt the teaching work to meet learning needs' (Black, Harrison, Lee, Marshall, & Wiliam, 2003, p. 2). Research into formative assessment suggests that it has the potential to produce significant learning gains (Black & Wiliam, 1998).

Muijs and Reynolds (2010) note that the aims of formative assessment may frequently be in tension with other conflicting uses and purposes of assessment in schools. These include: evaluating the quality of teaching and learning of individuals and whole school communities for various purposes; using this information to hold individuals to account for standards of teaching and learning; controlling the curriculum (via the ‘washback effect’ of high stakes examinations); and providing multiple stakeholders with assurances that an individual has reached a designated level of achievement.

Some of these conflicting purposes of assessment have been evident in MFL classrooms in England. In common with other school subjects, assessment of language learning has been shaped for many years by external assessment requirements: for example, the National Curriculum ‘levels’² at Key Stage 3; GCSE and A Level specifications at Key Stage 4 and sixth form respectively³. In this chapter, we focus particularly on Key Stage 3, where both the interventions we report were targeted. In this phase of education, the National Curriculum ‘levels’ were deeply embedded in schools’ practice, even though they were heavily criticised by SLA researchers for having a negative ‘washback’ effect on classroom practice. It was argued that they entrenched teaching methods that were not supported by research into how languages are learned (Macaro, Graham, & Woore, 2015).

The policy context surrounding language teaching and learning in England has undergone rapid change in recent years, and in relation to assessment at Key Stage 3 there have been several key shifts. First, the requirement to assess pupils’ progress in Key Stage 3 via the National Curriculum levels was abolished in 2013 and no alternative statutory framework has been put in its place. Second, alongside the emphasis on communicative approaches which had characterised the National Curriculum for MFL since its inception (DES/Welsh Office, 1991), the 2013 National Curriculum placed a greater emphasis than had previously been evident on what might be considered more ‘traditional’ aspects of language learning. These include explicit reference to grammar, translation and the study of ‘great literature’ and a

² The assessment framework for the National Curriculum in England, prior to the revisions in 2013, consisted of eight levels within each of one or more subject-specific attainment targets. For MFL there were four such attainment targets (speaking; listening and responding; writing; reading and responding).

³ Key Stage 3: students in their first three (sometimes two) years of secondary education, aged 11-14. Key Stage 4: students in the subsequent phase of education leading up to their compulsory national examinations, the GCSE (General Certificate in Secondary Education). Sixth form: two years of post-16 education leading to the Advanced Level (A Level) examination.

similar emphasis was to be seen in the revised subject content of GCSE and A Level examinations from 2015 onwards. Finally, these changes have been accompanied by a reduction in the availability of local and national support and advice structures previously accessible to MFL teachers, following the removal of most central funding for languages and the discontinuation of the National Languages Strategy (Johnstone, 2014).

All of these policy changes are recognised as having a significant impact on teachers' classroom practice (see Tinsley & Board, 2017). Our focus in this chapter is on two interventions that address specifically the removal of National Curriculum levels at Key Stage 3. As with all subjects, the lack of any prescribed assessment framework within the new National Curriculum has left behind a vacuum, resulting in some fragmentation of practice across contexts. This presents both opportunities and challenges for schools, with a clear need for the careful development and piloting of alternative assessment frameworks. In the case of MFL, this has posed particular problems for a number of reasons.

First, we would argue that the old NC levels were highly problematic in terms of the model of progression that they represented within each of the four language skills (Macaro, Graham, & Woore, 2015; Mitchell, 2003). Nonetheless, they cast a long shadow, continuing to exert a heavy influence on assessment practices in schools: for example, when we surveyed MFL departments in our own ITE partnership in 2016, twelve out of 16 departments were either still using the old NC assessment levels, or were using modified versions of them.

Second, assessing language proficiency is highly complex. One problem is that pupils may be stronger in one skill (e.g. listening) than another (e.g. reading); therefore, trying to sum up pupils' performance in a single overall grade or level is inherently reductive. Even a pupil's level of proficiency within a single skill is itself a complex construct. Different learners may achieve the same level of communicative competence through different balances of linguistic knowledge and strategic behaviour: for example, two pupils – Pupil A and Pupil B – may get the same mark on a reading comprehension exercise, but for different reasons. Pupil A may have understood most of the words but been unsettled by the small proportion that they did not know, leading them to miss important elements of the text's meaning. By contrast, Pupil B may have known relatively few of the words in the text, but been able to compensate for this lack by using sophisticated comprehension strategies. The fact that both pupils might achieve the same 'level' in an assessment illustrates that this judgment is of little formative

value in terms of helping them (and their teachers) know what they should do next in order to improve.

Third, SLA research has shown that progression in learners' linguistic output may follow a 'U-shaped' trajectory (McLaughlin, 1990), whereby they initially produce correct forms by reproducing memorized chunks of language (i.e. accuracy is high). As their knowledge develops, they may then begin to 'unpack' these chunks and generate language of their own, but thereby make more errors (i.e. accuracy decreases). Only later, when their knowledge of the language becomes more secure, do they return to producing correct forms (i.e. accuracy increases again to its previous levels). Some researchers have also identified a trade-off between complexity, accuracy and fluency, all of which are desirable features of language output and can be seen as different aspects of overall proficiency (see Housen & Kuiken, 2009). For example, learners may be able to produce language more accurately and fluently by restricting the complexity of what they are trying to say; but if they try to use more complex language, their fluency and accuracy may decrease. The fact that accuracy may actually decrease as learners progress is of course problematic for any assessment system which assumes that pupils will show a neat linear progression from one 'level' to the next.

Finally it is difficult to assess progression in MFL when (in our experience) schools' schemes of work often adopt a topic-based approach, in which the assessment at the end of each unit is highly dependent on the particular vocabulary associated with that topic. There is often little recycling of this vocabulary in subsequent topics, and little emphasis on the high-frequency words and structures that provide a basis for communication across topics.

Therefore, we can see that creating a valid and workable assessment framework in MFL is a complex matter indeed. Yet, since the removal of the NC levels in 2013, teachers in MFL departments across England have been faced with the considerable challenge of creating (or choosing) and implementing new assessment frameworks for their own particular contexts. In a sense, they can thus be considered to be conducting their own 'classroom interventions': they are making conscious changes in classroom practice aimed at improving learning and teaching. Anecdotal evidence, however, suggests that many teachers found themselves having to implement assessment frameworks in a hurried and sometimes unprincipled way – either producing something new within a very limited time-scale, adapting the familiar old levels, or buying in a commercial assessment scheme. Further, there is often a requirement

for these new assessment frameworks to comply with the constraints of generic frameworks which have been agreed at a whole-school level – even where teachers may consider these problematic for use in MFL classrooms specifically.

In response to these issues, the two studies that are the focus of this chapter take a more systematic and research-informed approach to investigating the effects of alternative models of assessment. The first was carried out by a practising MFL teacher who was also a student on a part-time Master's programme in Learning and Teaching. The study formed the basis for her final dissertation. The second study, led by university academics working in collaboration with student teachers and mentors in a range of schools, was an attempt to explore some of the possibilities of extending a multi-level school-university partnership (within the Oxford Education Deanery: see Fancourt, Edwards & Menter, 2015) beyond its traditional 'core' focus on Initial Teacher Education.

Intervention study 1 – Katherine's assessment study

Background to the study

The study in question was designed as an intervention to address a specific problem that had been identified in relation to both Katherine's own classroom teaching and the wider issues in MFL assessment raised by the reform of the National Curriculum in England, as outlined above. While the teacher had a general interest in developing her understanding of classroom-based assessment, she and her colleagues within the MFL department were also questioning more specifically the nature of the regular assessments that were being carried out in their school (every 6-7 weeks).

First, she was aware that, in the separate assessments for each of the four language skills (listening, speaking, reading and writing), pupils tended to perform at different levels for each, yet the score or 'level' awarded for each skill was aggregated to form one overall score for each pupil; she worried that this might not accurately reflect that pupil's actual skill-related performance. Second, she had concerns around the validity of the actual assessments being carried out, since assessment tasks for each skill were often presented in a way that made it difficult to determine an overall level of pupil competence. The mixing of different

levels of questions within the same assessment required the teacher to make a ‘best fit’ judgement as to the whether or not a pupil had achieved a particular level overall.

Furthermore, Katherine was responding to concerns shared with colleagues within the MFL department that they needed to be able to give pupils a more accurate picture of their progress. However, this had to be done whilst at the same time satisfying the school’s requirement for teachers to submit regular assessment data by which progress would be tracked at a whole-school level.

Finally, at a broader level, Katherine also wanted to develop her teaching in such a way as to incorporate more of the features of effective Assessment for Learning (AfL) within her classroom. Aware of the reported benefits of self-assessment and self-regulatory learning to both the learner and the learning process (Sadler, 1989; Assessment Reform Group, 2002; Swaffield, 2011; Black & Jones, 2006) she was keen to develop an assessment framework which included opportunities for her pupils to self-assess their own level of competence against specific criteria. An initial analysis of the pupils’ workbooks indicated that there was little effective self-assessment taking place prior to the intervention and that the pupils were very dependent on the teacher for any feedback on their work, with little opportunity to engage with the feedback they were receiving.

Thus, there were a number of starting points for the intervention. However, all were closely related to the wider context of the removal of statutory assessment levels and the subsequent onus placed on schools, and individual subject departments within schools, to devise their own assessment frameworks to measure pupil progress. Furthermore, the teacher was interested in exploring her pupils’ engagement with the process of learning a foreign language and felt that a more relevant assessment framework might lead to greater levels of engagement and motivation among some of her pupils. This view was influenced by Coleman’s (2009) review of motivation studies, which identified a lack of confidence and an inability to perceive their own progress as some of the key reasons for low motivation amongst MFL learners.

The nature of the intervention

The two key research questions were:

1. Can an assessment framework be developed for the MFL KS3 classroom which enables pupils to feel at ease with the assessment, while at the same time ensuring that their progress is being accurately measured?
2. Can self-assessment be used effectively within such a framework?

It was decided that a two cycle ‘action research’ approach would allow the issues to be investigated in a systematic and practical way (Cohen, Manion & Morrison, 2007; Elliott, 1991; McNiff, 2002). Further, as Edwards and Talbot point out, this style of research gives ‘scope for inventiveness and creativity’ (1997: p 61). Their ‘plan-act-monitor-review’ approach was used as the basis for both cycles of intervention (Figure 1).

(Insert Figure 1 here)

The first cycle saw the planning and implementing of an assessment approach that led to an increase in the amount of pupil self-assessment that was being carried out. This self-assessment was done against specific criteria devised by the teacher. The second phase developed these self-assessment criteria further, including the formulation and trialling of a more detailed and coherent framework for assessment. Here, pupils were able to assess their own performance against criteria that could be equated to ‘levels’ or ‘grades’ (for the purposes of whole-school monitoring purposes), yet which also enabled them to judge whether they were currently working ‘above’, ‘at’ or ‘below’ expected levels of performance. For each of these three levels of performance and within each separate skill area (listening, speaking, reading and writing), the pupils were given a series of ‘can do’ phrases with which to assess their own performance, and these were aligned to designated levels, enabling the teacher to track progression in each of the four skills individually over time.

The assessment framework that was developed and subsequently applied was informed by a reading of the relevant literature and drew on four existing frameworks: the previous National Curriculum level descriptors (Qualifications and Curriculum Authority, 2007); the GCSE grade descriptors from one of the national examination boards (AQA, 2014); the Languages Ladder grade descriptors (DCSF, 2007); and the MFL Professional Development Consortium’s assessment framework (PDC in MFL, 2013; see also below).

The participants

A class of 29 year 8 (12-13 year old) pupils learning Spanish was selected to be the focus of the study. This was a mixed-attainment class which Katherine had taught previously and one with which she felt she had a good relationship.

The first cycle of the intervention was conducted over one half-term, covering one unit from the MFL department's scheme of work; the second phase took place over the subsequent half-term, again covering one unit of work. A parallel class (taught by another teacher) acted as an informal 'comparison group'. This class was given the same assessment materials but they were not used regularly or explicitly in the lessons.

The first cycle focussed on the topic of 'holidays' and an assessment record was produced that would be presented in the form of a customised passport, in which the pupils would record their levels of understanding and progress. For the second cycle (focussed on the topic of 'leisure time and activities'), customised diaries were created to fulfil a similar purpose. Data were collected from both the intervention group and the comparison group through questionnaires and interviews, as well as through analysis of the passports and diaries themselves. There was a limited amount of collaboration within the MFL department while the study was being carried out.

The outcomes of the intervention

The intervention indicated that the framework had provided an opportunity to develop the learners' capacity for self-assessment and that it was also sufficiently rigorous to enable teachers to report progress at both the individual and whole class level. The pupil judgements were moderated by the teacher and her colleagues, and as the pupils became more familiar with the processes involved, there was evidence that they were able to make reliable assessments of their own performance against the designated criteria.

What was particularly interesting was the gradual development of a greater focus on learning as being the key to higher levels of performance. In the first survey, a greater proportion of pupils identified appropriate classroom behaviour as a factor that could lead to improvement (74%, with only 26% identifying a learning focus) but by the beginning of the second cycle, the proportions were 59% and 41% respectively. (In the comparison group, the second

survey revealed figures of 80% and 20% respectively). The data also indicated that lower attaining pupils tended to identify more targets related to their behaviour, whereas higher attaining pupils tended to identify learning targets.

The initial cycle revealed that there needed to be greater clarity in relation to the assessment criteria by which the pupils were being asked to assess themselves, and also the opportunity for greater levels of independence when doing so. While some pupils began to use the language of the success criteria when discussing their progress, the study also revealed that such engagement might nevertheless mask underlying misconceptions in terms of what the assessment criteria actually meant. One pupil summarised her achievements at the end of the intervention as follows:

'... in writing I can include some tenses and different vocab and overall I can cope with unfamiliar vocabulary in a topic I know'.

When asked, however, to give an example of this in practice, the pupil replied:

'What's vocabulary? I don't know what it is.'

One other interesting outcome was in relation to the pupils' perception of the subject itself. Prior to the intervention many of the pupils were generally motivated to learn but nevertheless lacked confidence in their ability to perform in one or more language learning skills. There was some evidence that, in spite of continuing to find the subject challenging, pupils felt more confident in attempting to use the language productively, with some pupils commenting on the value of the self-assessment grids contained within the passports and the diaries, for example:

'It does give you a little bit of hints like saying what you need to do or what you have been doing so it just helps you in a sense and if you are in the middle and you look at the top and you realise you haven't done that then it's kind of helped you, not only to realise what you're meeting but also how to improve.'

Dissemination of the results of the intervention

The teacher was able to disseminate her findings both within her school and beyond, since she moved to a teaching post at a different school the following academic year, where she discussed the study and its findings with her new colleagues. In addition, there was the opportunity to lead an ITE seminar at the university through which she was completing her Master's course, focussing on a number of assessment issues (including the validity and reliability of assessment data) with beginning teachers.

Case Study 2: trialling a 'Pedagogical Assessment Framework'

Background to the study

The second study was one in which student teachers on an Initial Teacher Education course trialled an alternative assessment framework in MFL, the 'Pedagogical Assessment Framework' or 'PAF' (Macaro, Graham & Woore, 2015). This framework was developed as part of the 'Professional Development Consortium in MFL' (pdcinmfl.com), a collaborative project involving secondary school teachers and university-based researchers. Over a period of six weeks, working with the support of their school-based mentors and university tutors (the current authors and other colleagues), the student teachers each used this framework to assess a specific language skill (either reading or speaking) with at least one class in Key Stage 3.

A brief description of the PAF is in order at this point. (A full copy of the framework can be found at pdcinmfl.com and in Macaro et al., 2015). At its core is a twin approach to assessment for each language skill, comprising a 'main strand' (focussing on communicative competence) and a 'supporting strand' (focussing on the underpinning linguistic knowledge – e.g. knowledge of vocabulary and grammar – and strategic behaviour). The purpose of the main strand is to ensure that the principal focus is always on what pupils can actually do with the language. The purpose of the supporting strand is to provide diagnostic information to inform teaching and learning: it helps learners (and their teachers) decide on the priority areas to work on in order to develop their communicative competence further. To illustrate, Pupil A (in our example above) would benefit from developing their strategic competence, helping them to be more resilient when faced with unknown words. Pupil B, by contrast, is already

an effective strategic reader, but this will only carry them so far: they must increase their vocabulary knowledge in order to progress to more complex texts.

Again, the launch of the new NC and its lack of any statutory assessment framework provided the immediate impetus for this study. The resulting fragmentation of assessment practices across partner schools presented a practical problem in relation to how the coverage of assessment on the ITE course would be addressed. Discussions of this issue in university seminars could no longer build upon a foundation of common practice in schools. Indeed, several student teachers raised this issue in end-of-course evaluations as an impediment to their progress on the course. Furthermore, informal observations and conversations with teachers in partner schools indicated that MFL departments were themselves grappling with the implications of the new NC and the urgent question of how to handle assessment in the ‘post-levels’ context.

These considerations led to the idea of an intervention study which could: develop the student teachers’ understanding of assessment; develop assessment practices in schools across the partnership; and more broadly explore the viability and effectiveness of the PAF framework itself, including its washback effects on classroom teaching and learning. The project sought to address a number of questions but the ones focussed on in this chapter are as follows:

1. What were the perceived strengths and weaknesses of the PAF?
2. What effects did it have on classroom teaching and learning?
3. What was the impact of the project on the student teachers’ understanding of assessment?

Nature of the intervention

The project can be conceptualized as a ‘dual’ intervention operating on two levels. First, a change in practice in the ITE course was instigated, resulting in the student teachers being asked to try out the PAF with one of their classes, and being supported in doing this. Second, the student teachers’ use of the PAF was itself an intervention in their own classrooms, since they were trying out a different assessment framework from the one otherwise in use in their school context.

For the first (ITE-based) intervention, it was considered important that the student teachers were not simply presented with the new assessment framework and a set of instructions to follow in order to implement it. Rather, university sessions were used help them to understand the PAF and its underlying rationale. It was hoped that they would then be empowered to use it in ways with which they felt comfortable and which fitted their own particular classroom contexts. In other words, they were able to take some degree of ownership over how the assessment framework was to be implemented⁴. For the second (school-based) intervention, student teachers were allocated randomly to one of two groups, the first of which focussed on speaking and the other on reading. They then used the PAF to assess the relevant skill over one school half-term (March/April).

The school-based intervention had a number of limitations: for example, student teachers were being asked to experiment with a new system of assessment when they themselves were novice in matters of assessment generally. The intervention was also shorter than would have ideally been the case, being constrained by the ITE course calendar. Further, in light of the short time scale and the student teachers' status in school, it was not possible to embed the PAF as the 'official' assessment framework (the one used to communicate with parents and school managers about pupils' progress). These factors limited the effects of the intervention but nevertheless allowed a flavour of its impact to be obtained.

Methods of data collection

From the cohort of 28 student teachers, 23 participated in the project, spread across 15 schools. 12 focussed on reading and 11 on speaking. They were supported by 15 school-based mentors and five university tutors (including three of the present authors). Three schools opted out of the project, since they did not feel able to support their student teachers' use of the PAF when they were already in the process of implementing their own new assessment frameworks. This accounts for the five student teachers who did not participate.

Data were gathered from a range of sources. Both the student teachers and their mentors completed on-line questionnaires eliciting their views on the PAF (and on their participation

⁴ Student teachers had access to a range of exemplar assessment tasks that might facilitate the implementation of the PAF. They used their own judgement to modify and/or re-design these tasks to suit the work being undertaken in their own classrooms.

in the project itself), and follow-up interviews were conducted with the student teachers. Samples of assessment materials which the student teachers created (e.g. reading comprehension tasks, speaking tasks and vocabulary tests) were also collected. This was to provide insight into how the framework was implemented in practice: i.e. how the student teachers translated the assessment framework into concrete assessment tasks. Finally, samples of pupils' work produced in response to these tasks were collected, together with any feedback provided by the student teachers. The findings presented in this chapter are based on the questionnaire and interview data.

The data collection (and indeed the project as a whole) presented some complex ethical issues relating to the power relations between the different groups of participants. Participation was optional for both the student teachers and their mentors, but given that the university tutors were not only the researchers, but also the course tutors, there was a risk that the student teachers might feel coerced into taking part. To guard against this, it was repeatedly stressed that their participation in the project was entirely unconnected with their own progress and assessment on the ITE course. Further, a research assistant was employed to manage and conduct the data collection, thus making it easier for the student teachers to opt out should they wish to do so. This mechanism was also designed to increase the likelihood of the student teachers giving their honest opinions on the PAF, rather than saying what they thought their tutors might want to hear.

A further tier of complexity was created by the mentors' involvement in the project. There was significant variability in the degree to which mentors were prepared to relinquish control of assessment procedures their classrooms, in order to enable student teachers to experiment with the PAF.

Outcomes

Here, we provide a brief overview of some of the project's key findings; see Woore, Mutton, Molway, Macaro and Savory (in preparation) for a more complete account of the outcomes.

In relation to the first research question, the student teachers mentioned various weaknesses of the PAF, but these were disparate, with no particular patterns emerging. For example, they felt that: the PAF was time consuming to implement (n=2 respondents); it was unclear how to

interpret some of the terms in the level descriptors, such as ‘familiar language’ (n=2); the levels were pitched too high for beginner learners (n=1) or were too broad to allow their pupils’ progress to be documented effectively (n=1); and the label ‘beginner’ attached to the first level could be demotivating (n=1). Further, two participants mentioned difficulties with reliability, as illustrated by the following comment made in one of the interviews:

I know that there is a lot of criticism about the previous system [the old NC levels] (...) you tick this tense and you go up a level (...) However it does give you some clear criteria from the teacher’s point of view and the students’ perspective, when they are trying to complete tasks, when they are aiming for something.

This response is interesting, however, because it seems to acknowledge that the (perceived) greater reliability of the old system came at the expense of its validity as a model of progression in foreign language learning.

By contrast, there was clear agreement on a number of strengths of the PAF. The responses of 15 out of the 23 participants indicated that, in broad terms, they felt the PAF offered a valid model of progression in language learning, and was a useful diagnostic tool for understanding learners’ progress and current learning needs; this understanding, in turn, could inform subsequent lesson planning. By contrast, the old NC levels (still in use in numerous partner schools at that time) were felt to embody an incomplete and inaccurate model of progression, placing artificial emphasis on particular aspects of language learning whilst neglecting other important aspects. The following quotations from two participants illustrate these views:

To me, what is best about the [PAF] framework is that it makes teachers think about what the exact building blocks for success in language learning are, as these are very different from what has been conveyed in the old NC levels.

I found the PAF really useful to work out what I needed to do to help improve the students. For example, I could address their overuse of cognates and give the students strategies. It gives me more information about the students’ needs. The [old NC] levels in comparison are superficial. I would say the PAF provides a good diagnosis for future planning.

To address the second research question, the student teachers were asked whether (and if so, in what ways) their use of the PAF had influenced their wider classroom practice. Thirteen participants responded to this question. Two of these reported little influence on their classroom practice, because they had had insufficient time to engage with the PAF as thoroughly as they would have liked. One said that there had been a negative influence, because they felt that the PAF had compromised their ability to cover the school's prescribed scheme of work. However, ten respondents felt that the PAF had had a positive impact on their teaching. Specifically, they reported that they had put more emphasis on the following, when compared to the teaching they had been doing previously (before they started using the PAF): developing pupils' language learning strategies (n=4 participants); their knowledge of high frequency words (n=2); their proficiency in phonological decoding (n=1); and their ability to engage in spontaneous oral interaction (n=1). Another said that the PAF had led her to focus more instructional attention on the *processes* of language learning, rather than simply the product. It could be argued that all of these reported outcomes are highly positive ones, when seen from the perspective of research into language learning and teaching (for an overview see Macaro, Graham & Woore, 2015).

Five other student teachers, when asked about the influence of the PAF on their classroom teaching, reported that it had developed their understanding of the underlying components of language proficiency, of how pupils make progress in these areas and hence of how this might be reflected in their teaching. As one respondent noted:

It made me think how best I could prepare students for an assessment based on those levels – just like when teaching for national curriculum level-based assessments, you teach students two tenses so they can get a level 5⁵. It was much more positive than the latter situation, however, because I was more concerned with giving students the tools and practice they needed to successfully communicate, rather than focussing on being able to reproduce certain structures just because someone somewhere was dictating that they should.

⁵ The nature of the previous National Curriculum descriptor for level 5 in writing included the requirement that pupils should 'refer to recent experiences or future plans, as well as to everyday activities'. This became interpreted by many teachers as being an indication that the pupils had to produce at least two different tenses within one assessed piece of work in order to attain a level 5.

Finally, in terms of impact of the project on the student teachers' own professional learning (research question 3), it is clear from analysis of their questionnaire and interview responses that they had developed a deeper and more critical understanding of the issues involved in assessing pupils' progress in MFL. This is evident, for example, in the four quotations above, in which the student teachers discuss the validity of the PAF in relation to the old NC levels, backing up their views with specific arguments relating to the nature of foreign language learning. Further, as had been hoped, the PAF provided tutors with a common point of reference when discussing assessment-related issues in university seminars.

Dissemination

The findings of the PAF project have been (or are being) disseminated by the university-based authors via various conference presentations and published outputs, as well as through local school-university partnerships.

It could, however, be argued that the intensive nature of the work undertaken with the student teachers as part of this project – for example, the detailed critical discussions of the PAF, its background and the issues arising in its implementation – may well have had a deep and potentially long-lasting influence on their classroom practice. Of course, this influence may extend, in the immediate term, only to the classrooms of those student teachers who participated in the project, or of those working directly with them in schools. However, this tension between 'breadth' and 'depth' of impact is something that, we would argue, merits further discussion.

Discussion

These two small-scale, mediated classroom interventions raise a number of interesting issues which potentially have a wider relevance beyond the MFL contexts in which the studies were carried out. These issues can be summarised in terms of three specific questions.

1. *What have we learnt about classroom-based interventions in assessment and the related challenges?*

Both interventions presented in the above case studies focus on the development of assessment practices in MFL classrooms following the removal of the statutory and well-established National Curriculum levels at key stage 3. In comparison to a more straightforward intervention that introduces, for example, a specific new teaching strategy over a period of time (the impact of which might be measured from data gathered through pre- and post-tests), both the interventions described above are complex. The process of devising and applying alternative assessment frameworks and then creating related assessment tasks is sufficiently complex, but then to try to evaluate the validity and reliability of these frameworks increases this level of complexity even further. In the PAF intervention, there was some attempt to ‘triangulate’ the student teachers’ assessment judgements by asking the mentors working directly with them to moderate the assessments made. In Katherine’s case, where the focus was also on developing the pupils’ capacity to self-assess their language competence accurately, she used her own assessment to moderate the pupil judgements. In both cases, however, a certain degree of subjectivity was involved, raising questions as to the reliability of the approaches taken beyond the specific contexts in which the interventions were carried out.

A further factor in determining possible limitations of the two interventions relates to the validity of the assessment criteria that were developed in both studies. The validity of the assessment criteria within any given framework can only be evaluated through the fluent use of the framework and its criteria within the classroom, but such use presupposes pupil familiarity with the criteria and the ability of the teacher – and the pupils - to engage with them effectively. This, in turn, requires an initial process of introduction and familiarisation. However, to do this, more time and resources would have been necessary than were available to either of the interventions described in this chapter.

The need for a shared understanding of assessment criteria is illustrated clearly in Katherine’s study where one of the pupils *appeared* to be able to carry out the process of self-assessment effectively and to highlight areas which would serve as his future learning targets; in doing so he was able to articulate the language of the success criteria as expected, but demonstrated little understanding of the meaning of quite basic terminology used within those criteria (that is to say the meaning of the word ‘vocabulary’). Questions are thus raised about the common level of understanding and interpretation of the criteria. Nevertheless, it could be argued that Katherine was only able to gain this greater insight into the pupil’s lack of understanding

because of the detailed data collection that accompanied the action research project. This misconception might otherwise have remained hidden.

Overall, whilst we recognize that assessment in any subject area is complex, we would argue that this is particularly so in MFL, where the principal aim of instruction is the development not of conceptual understanding, but of communicative competence. This complexity is further exacerbated by the pressure to satisfy institutional requirements to produce regular assessment data within a prescribed format for accountability purposes. Drawing on research into the way in which languages are learnt can, however, help us to devise assessment frameworks that are both valid and promote effective pedagogical approaches. In both the interventions described in this chapter, the focus is ultimately on learning and teaching, rather than on the wider issues around assessment policy and accountability issues, leading to a more positive ‘washback’ effect on classroom practice.

2. *What have we learnt about the ways in which interventions of this nature contribute to teacher professional development and learning, and the related challenges?*

Professional autonomy

One key issue relates to the level of individual teacher autonomy within classroom intervention studies of this nature. In Katherine’s case, although constrained by the requirement to submit regular assessments of her pupils’ progress to senior leaders in the school, she was nevertheless able to take the initiative to devise a way in which such data could be generated through a more innovative approach. She was able both to create a new assessment framework (based on her reading of the literature and informed by her understanding of other such frameworks) and to create the tasks by which pupil progress would be measured, drawing on her own interests and capabilities (Benson, 2010).

In the case of the PAF, the student teachers did not have a hand in developing the assessment framework itself, and they also had to use the PAF within the constraints of the existing policies and practices of the schools in which they had been placed. Nonetheless, they were able themselves to devise assessment tasks which they considered appropriate for the specific contexts in which they were working, which again allowed them to make the most of their own strengths and to develop their own ideas. The student teachers involved in the PAF

project told us that the structured nature of the enquiry and the expectations around such involvement afforded them the necessary ‘permission’ to develop greater levels of agency than might otherwise have been the case.

Professional knowledge base

Further, the PAF study highlighted the need for teachers to have the necessary systematic knowledge and understanding to be able to challenge, where appropriate, existing policies and practices and to be equipped to deal with change effectively as part of the ongoing process of continuing professional development. As noted above, a number of colleagues within our partner schools found it challenging to respond to the sudden removal of the previous National Curriculum levels and the requirement to create (or choose) a new assessment framework at short notice. We would argue that both Katherine and the student teachers who took part in the PAF study may have been better equipped to deal with such a challenge, since they had been required to draw on a range of perspectives in order to enable them to judge the validity and reliability of specific assessment approaches that they were undertaking in their classrooms. Thus teachers’ work is seen as a ‘professional endeavour’ in which:

their reflective abilities should bring together their own experience and the deliverances of research to enable them to determine both short and long-term courses of action.

(Winch, Oancea & Orchard, 2015: pp. 210-211).

Likewise, Katherine was able to address the challenges of introducing a new assessment framework into her classroom practice because she was able to draw on a range of different theoretical perspectives before considering what might work within her specific context. Her reading of the wider assessment research literature and her analysis of existing frameworks enabled her both to develop an alternative set of assessment criteria (which would provide the data needed to satisfy the school’s requirements for regular assessment updates) and to introduce key aspects of formative assessment into her classroom, particularly the use of pupil self-assessment.

Professional learning

In spite of such calls for research to underpin teachers' practical work in the classroom, the extent to which teachers are able to engage both with and in research (BERA, 2014) does raise important questions in relation to Initial Teacher Education and ongoing professional development. One such question concerns how this 'research literacy' (BERA, 2014, p.7) can be achieved. A key factor appears to be the involvement of specialist expertise (likely to be external to the school) which is able to illustrate innovative approaches, assist with the process of evaluation and scaffold the professional learning around these approaches (Cordingley, 2015). In both the studies examined in this chapter, there was external support in place, albeit from two different perspectives. In Katherine's case she was, as a Master's student, supported in her intervention research through regular discussion with and feedback from a university supervisor; in the case of the PAF project, the intervention was set up by the university, which provided both the initial grounding in the knowledge and understanding that might be required (delivered as part of the normal ITE programme) and ongoing support through individual monitoring by course tutors.

Similarly Cordingley (2015) identifies peer support as one of the key contributors to effective professional learning, through 'collaboration, especially reciprocal risk taking and professional dialogue' (2015: p. 240). The MFL interventions described here indicate both the strengths and the limitations of such collaboration. Although Katherine had the support of her colleagues in the school, the intervention was carried out very much within her own classroom and with little wider professional discussion. Colleagues had the opportunity to be involved but appeared to have little incentive to do so, in spite of the concerns that had been expressed collectively in relation to the schools' current practices for assessing and recording pupils' progress. Teachers may be sceptical about the value of educational research and not see its relevance to practice (Vanderlinde & van Braak, 2010), yet interventions of the sort that Katherine carried out are driven by a sense of immediacy and are carried out in real classrooms in which the teachers in question have a detailed understanding of the specific contextual factors involved. It may nevertheless still be the case that colleagues have insufficient interest in trialling innovative approaches and may prefer instead to implement approaches prescribed from elsewhere.

Where the PAF study differed was that it was deliberately set up in such a way as to require collaboration between a range of different research participants (university researchers, school-based mentors and student teachers). The fact that the interventions were being

carried out on multiple sites enabled discussion to be focussed on both the enquiry process and the emergent findings. The student teachers had a high level of responsibility for introducing innovative approaches, albeit within limited parameters, yet did have the necessary support structures in place to enable them to seek guidance where necessary, get feedback on the assessment tasks they were devising and share experiences of the intervention with others. We would therefore argue that the ambition to develop research literacy among teachers (BERA, 2014) is more likely to be realised if opportunities to do so are structured, rather than simply left to chance.

3. What have we learnt about the interventions themselves?

Scale and relevance

A further general issue raised by the studies reported in this chapter concerns their scope and their potential for wider impact. To what extent can such small-scale research produce knowledge with relevance beyond the specific classrooms and contexts in which it takes place? In both cases, it could be argued that there was real ‘depth’ of impact (albeit on a small scale). The interventions arguably led to a number of potentially beneficial outcomes related both to the participants’ understanding and practice of assessment and to the pedagogical benefits of adopting new approaches to classroom-based assessment. The latter included a greater focus on the diagnostic aspects of assessment (Black & Wiliam, 2009) and the promotion of more pupil self-assessment in order both to enhance the learning process and to increase pupil motivation (Black, Harrison, Lee, Marshall & Wiliam, 2004). Readers may note that both our studies lack the measurement of pupil outcomes. However, this would have been difficult given that the focus of the studies was, precisely, the development of valid ways to measure pupil outcomes.

In terms of the assessment frameworks themselves, these were tested and, within the limitations of the individual interventions, judged to have some validity. Although these were small-scale, exploratory interventions, we would argue that they nonetheless represent a valuable first step in a wider process of the systematic, research-informed development of a new assessment framework. This process is particularly important in a context in which some schools may be feeling pressurized into making hurried innovations in their assessment

practices, which risks the adoption of less principled frameworks or ones which are less appropriate in an MFL context.

However, the full impact of small-scale interventions as a first step can be realized only if their findings are available to other researchers and practitioners. Here we would note an imbalance in the possibilities for such dissemination between our two interventions. These possibilities tend to be more limited for classroom practitioners than for academic researchers, for whom publication and dissemination is already a professional expectation. We would therefore raise the question as to how the findings of such small-scale, mediated interventions might be disseminated more widely, given that many practitioners involved in such studies may produce interesting results, yet may not be in a position to communicate these beyond their immediate communities of practice.

Conclusions

The assessment frameworks trialled in our two interventions are attempts to develop subject-specific models of assessment which reflect research evidence concerning how languages are learnt, have positive washback on pedagogy and can play an effective formative role in guiding teaching and learning. Both frameworks were positively evaluated, at least within the limitations of these small-scale studies. These initial findings now invite further research to develop and evaluate the frameworks further, for the benefit of classroom practitioners. Thus, despite their exploratory nature and small scale, our studies represent potentially important early steps in a wider process of systematically developing and evaluating pedagogical innovations.

Beyond these specific findings, our studies also raise a number of broader issues relating to issues such as the relationship between research and practice; the potential impact of small-scale intervention studies; and the nature of teachers' professional learning. We address each of these points in turn.

First, we noted above that the relevance of educational research to classroom practice is sometimes viewed with scepticism (even hostility). However, both our studies illustrate how a research-informed approach can be effectively brought to bear on practical classroom problems. The immediate catalyst for the studies was the sudden shift in policy which removed the previous statutory assessment framework. However, the studies are also highly

attuned to the specific concerns of the particular contexts in which they were conducted. This, we argue, reflects their ‘mediated’ nature. As we noted above, although this was achieved in different ways in the two studies, both were deeply rooted in both theoretical and practical perspectives, with practical concerns being mediated by research understandings, and vice versa.

Second, we would highlight that both of the studies reported in this chapter have had considerable impact in the immediate contexts (schools and ITE partnership) in which they were conducted. However, as with any research, the extent to which their findings can have wider impact depends on the effectiveness with which these are disseminated. Currently (notwithstanding this chapter), practitioner research has tended to have less easy pathways to dissemination than that conducted by ‘academic’ researchers (such as university lecturers). Our two case studies are no exception. We would therefore like to see increased possibilities for dissemination of practitioner research, such that this becomes normalized as part of the teacher’s role as a ‘research literate’ professional. Of course, there may be differences in the methodological and theoretical rigour within this widened stock of published studies; however, this is something that individual teachers would continue to assess as part of their critical engagement with published research (Winch et al., 2015).

Finally, the contribution made by classroom interventions to teachers’ wider professional learning should not be underestimated. Indeed, this may afford significant additional benefits alongside the new knowledge arising from the interventions themselves. For example, both Katherine and the student teachers involved in the PAF project developed rich and sophisticated understandings of the issues involved in assessing progress in MFL, and of the relationship between assessment on the one hand and teaching and learning on the other. Here we see a real ‘depth’ of impact of the studies – in terms of changes to teachers’ beliefs and classroom practices – which contrasts with the more usual concern with ‘breadth’ of impact.

A question which then arises is to what extent these kinds of approaches to developing teachers’ research-informed understanding of pedagogical issues are ‘scalable’. Should all teachers have the opportunity to engage with and in research to the same depth as was achieved by Katherine through her Master’s course and by our student teachers through their involvement in the collaborative research project? Echoing the ambition of the BERA/RSA

(2014) inquiry, we would offer an affirmative answer to this question. Small-scale, appropriately mediated classroom interventions have the potential to play an important role in facilitating teachers' engagement with and in research.

References

Assessment and Qualifications Alliance (2014). *GCSE Specification Spanish*. Retrieved from: <http://filestore.aqa.org.uk/subjects/AQA-4695-W-SP-14.PDF>.

Assessment Reform Group (2002). *Assessment for Learning: 10 Principles*. Retrieved from: <https://www.aaia.org.uk/content/uploads/2010/06/Assessment-for-Learning-10-principles.pdf> .

Benson, P. (2010). Teacher education and teacher autonomy: Creating spaces for experimentation in secondary school English language teaching. *Language Teaching Research*, 14(3), 259-275.

BERA (2014). Research and the Teaching Profession: building the capacity for a self-improving education system. Final report of the BERA-RSA Inquiry into the role of research in teacher education. Retrieved from: <https://www.bera.ac.uk/wp-content/uploads/2013/12/BERA-RSA-Research-Teaching-Profession-FULL-REPORT-for-web.pdf>.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for Learning: Putting It Into Practice*. Maidenhead: Open University Press.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 8-21.

Black, P. & Jones, J. (2006). Formative assessment and the learning and teaching of MFL: sharing the language learning road map with the learners, *The Language Learning Journal*, 34, 4-9.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, 5(1), 7-74.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5.

Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics*, 32, 83-110.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education*, 6th Edition, London: RoutledgeFalmer.

Coleman, J. (2009). Why the British do not learn languages: myths and motivation in the United Kingdom, *The Language Learning Journal*, 37, (1), 111-127.

Cordingley, P. (2015). The contribution of research to teachers' professional learning and development. *Oxford Review of Education*, 41(2), 234-252.

Department for Children, Schools & Families (2007). *The Languages Ladder: Steps to Success*. Retrieved from:
<http://webarchive.nationalarchives.gov.uk/20130401151715/http://education.gov.uk/publications/eorderingdownload/dcsf-00811-2007.pdf> .

Department for Education (DfE) (2013). National Curriculum. Retrieved from:
<https://www.gov.uk/government/collections/national-curriculum>.

Department for Education and Science (DES)/Welsh Office (1991). *Modern Foreign Languages in the National Curriculum*, Department of Education and Science and the Welsh Office.

Edwards, A. & Talbot, R. (1997). Action Research *The hard-pressed researcher: a research handbook for the caring professions*. England: Addison Welsey Longman.

Elliot, J. (1991). *Action research for educational change*. Milton Keynes: Open University Press.

Fancourt, N., Edwards, A., & Menter, I. (2015). Reimagining a School–University Partnership: The Development of the Oxford Education Deanery Narrative. *Education Inquiry*, 6(3). 27724, DOI: 10.3402/edui.v6.27724.

Graham, S. & Macaro, E. (2008). ‘Strategy Instruction in Listening for Lower-Intermediate Learners of French’ in *Language Learning*, 58(4), 747-783.

Hassan, X., Macaro, E., Mason, D., Nye, G., Smith, P. & Vanderplank, R. (2005). *Strategy training in language learning - A systematic review of available research*. EPPI-Centre, Social Science Research Unit, Institute of Education, University of London, London. Retrieved from <http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=296> .

Housen, A. and Kuiken, F. (2009). ‘Complexity, Accuracy, and Fluency in Second Language Acquisition’, in *Applied Linguistics*, 30(4), 461-473.

Johnstone, R. (2014). Languages over the past 40 years: Does history repeat itself? In P. Driscoll, E. Macaro, & A. Swarbrick (Eds.), *Debates in Modern Languages Education*. Abingdon: Routledge.

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, 46(1), 1-40.

Macaro, E. & Erler, L. (2008). Raising the Achievement of Young-beginner Readers of French through Strategy Instruction, *Applied Linguistics*, 29(1), 90-119.

Macaro, E., Graham, S., & Woore, R. (2015). *Improving foreign language teaching: Towards a research-based curriculum and pedagogy*. London: Routledge.

McLaughlin, B. (1990). Restructuring. *Applied Linguistics*, 11(2), 113–128.

McNiff, J. (2002). *Action Research: Principles and Practice* London: Routledge

Mitchell, R. (2003). Rethinking the concept of progression in the National Curriculum for Modern Foreign Languages: a research perspective. *Language Learning Journal*, 27(1), 15-23.

Muijs, D., & Reynolds, D. (2010). *Effective teaching: Evidence and practice*. London: Sage. Professional Development Consortium in Modern Foreign Languages (PDC in MFL)(2013) *Alternative assessment framework for languages*. Retrieved from: <https://pdcinmfl.com/alternative-assessment-framework-for-languages>.

Oancea, A., & Pring, R. (2008). The importance of being thorough: On systematic accumulations of “what works” in education research. *Journal of Philosophy of Education*, 42(1), 16–39.

Plonsky, L. (2011). The effectiveness of second language strategy instruction: a metaanalysis. *Language learning*, 61(4), 993-1038.

Qualifications and Curriculum Authority (2007). *Modern foreign languages: Programme of study for key stage 3 and attainment targets*. Retrieved from: http://webarchive.nationalarchives.gov.uk/20110223175304/http://curriculum.qcda.gov.uk/uploads/QCA-07-3340-p_MFL_KS3_tcm8-405.pdf.

Sadler, D. (1989). Formative assessment and the design of instructional systems, *Instructional Science* 18, 119-144.

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasiexperimental intervention studies. *TESOL Quarterly*, 46(4), 842-854.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language teaching research*, 12(3), 329-363.

Swaffield, S. (2011). Getting to the heart of authentic Assessment for Learning, *Assessment in Education: Principles, Policy & Practice*, 18, (4), 433-449.

Tinsley, T., & Board, K. (2017). *Language Trends 2016/17: Language Teaching in Primary and Secondary Schools in England*. British Council.

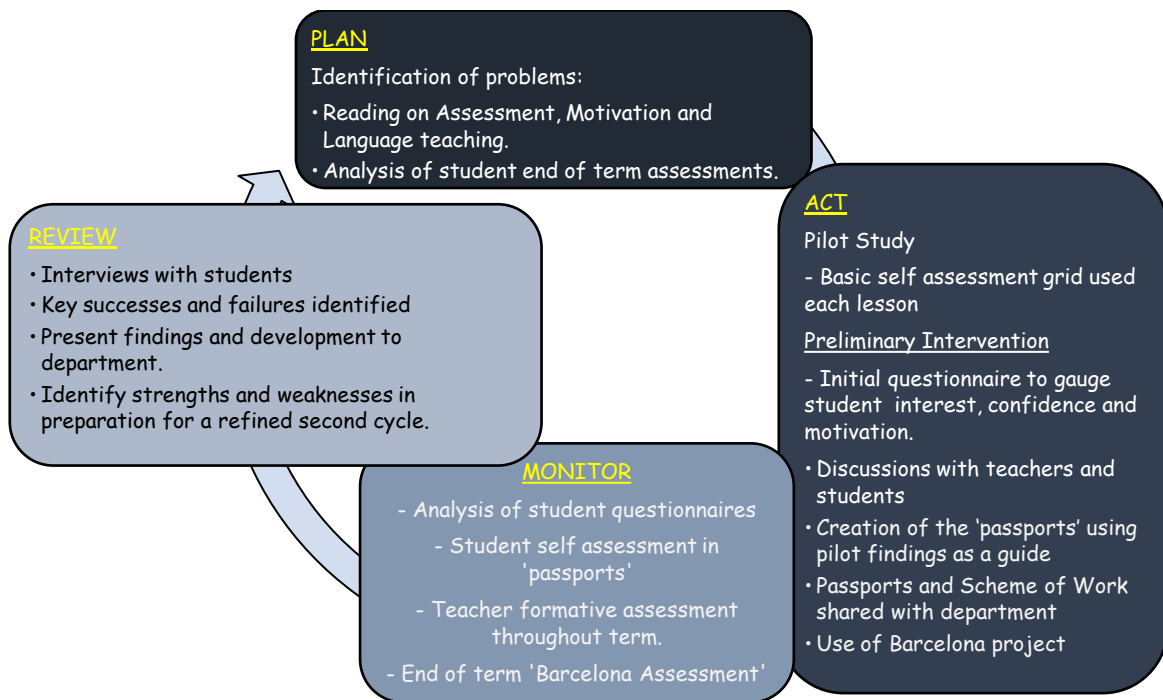
Thomas, G. (2016). After the Gold Rush: Questioning the “Gold Standard” and Reappraising the Status of Experiment and Randomized Controlled Trials in Education. *Harvard Educational Review*, 86(3), 390-411.

Vanderlinde, R., & van Braak, J. (2010). The gap between educational research and practice: views of teachers, school leaders, intermediaries and researchers. *British Educational Research Journal*, 36(2), 299-316.

Winch, C., Oancea, A., & Orchard, J. (2015). The contribution of educational research to teachers’ professional learning: philosophical understandings. *Oxford Review of Education*, 41(2), 202-216.

Woore, R., Mutton, T., Molway, L., Macaro, E. and Savory, C. (in preparation). ‘Towards a ‘Pedagogical Assessment Framework’ in Modern Foreign Languages: an attempt to align language teaching, learning, assessment and research evidence’.

Figure 1: overview of the action research approach



Cycle Two

PLAN

- Using strengths and weaknesses of the 'passport', creation of 'diary'.
- Further reading on assessment
- Consideration of GCSE descriptors, Level descriptors, PDC descriptors and Language Ladder
- Creation of progress grid as a synergy of the grade descriptors.

ACT

- Repeat of questionnaire to gauge student interest, confidence and motivation.
- Comparison group completes questionnaire
- Discussions with teachers in department of assessment grading
- Implementation of the 'diaries' with newly formed group.
- Use of progress grids in class through diary

MONITOR

- Analysis of student questionnaires
- Student self assessment in 'diaries' and progress grids
- Teacher formative assessment throughout term.
- End of term 'Diary Assessment'

REVIEW

- Interviews with students - from original class and comparison group.
- Key successes and failures identified
- Present findings and development of project to department
- Study of student progress grids in relation to actual performance
- Consideration of findings and implications