The Assessment of Welsh as a Second Language

A report on current Welsh language assessment practices in mainstream English medium primary schools

The Assessment of Welsh as a Second Language:

A report on the current Welsh language assessment practices in English medium primary schools.

Author: Peter Owen Russell

Research supervisors: Dr Sion Jones (Senior Lecturer, Cardiff University School of Social Sciences) and Dr Katy Jones (Senior Lecturer, Cardiff University School of English, Communication and Philosophy).

This document was generated as part of a doctoral thesis through Cardiff University's doctoral programme and the WGSSS.

Published: June 2025

Table of Contents

Summary of Acronyms

AfL – Assessment for Learning

CEFR – Common European Framework of Reference for Languages

CfW – Curriculum for Wales

CLIL – Content and Language Integrated Learning

CPD – Continuing Professional Development

CSC – Central South Consortium

EAS – Educational Achievement Service (Consortium for Southeast Wales)

EM – English Medium

EPI – Extensive Processing Instruction

ITE – Initial Teacher Education

MAT – More Able and Talented

SLA – Second Language Acquisition

SLT – Senior Leadership Team

WG – Welsh Government

WLC – Welsh Language Continuum

WLS – Welsh Language Sabbatical

WM – Welsh Medium

1. Introduction

This report includes the findings of a scoping study carried out between March and May 2025 exploring current Welsh language assessment practices in mainstream English medium (EM) primary schools. The study included details of the assessment protocols of 81 schools representing every region and consortium of the country. Data was also gathered on teacher attitudes to the assessment of Welsh within their institution and opportunities were provided for them to propose adaptations and augmentations that they felt would improve assessment and enhance learning in their institutional context.

This report begins with an executive summary of the key findings from the study. This is followed by a more thorough presentation of the study methodology, data, and the analysis carried out to identify these themes.

1.1 Research questions

The study aimed to collect data in response to four research questions:

- o What are the current formative and summative Welsh language assessment practices and tools being used in mainstream EM primary schools across different regions of Wales?
- o To what extent do teachers feel current assessment practices and provisions meet the needs of teachers and pupils in Welsh language learning?
- o How is assessment of Welsh language proficiency used as part of transition plans and processes to ensure continuity of learning?
- o What do teachers feel could be done to enhance Welsh language assessment provision in their school context?

2. Executive Summary of the Findings

The key findings of the report fall into five primary themes, covering the assessment tools provided by schools (both formative and summative), the formative assessment practices of teachers, the way assessment facilitates transition, the attitudes of teachers to assessment, and the recommendations made to enhance assessment of Welsh.

2.1 Current assessment tools

- o The study revealed that few schools provide teachers with any tools to support the assessment of Welsh. Only 13% of respondents reported being provided with provision for formative assessment, and 23% for the summative assessment of Welsh.
- o 60% of teachers who were provided with a formative assessment tool felt that the tool was inaccurate in measure learner progression. Overall, 5% of participants felt they had an assessment tool that was suitable for the measurement of Welsh competency in their school.
- o Of the 5% of teachers who felt they had a suitable formative assessment tool, all but one reported that this tool was unique to their institution. As assessment tools were not shared by other schools in the cluster, it did not provide a sense of inter-institutional learner competence. This was supported by the low proportion (23%) of teachers who felt they had a moderate or better understanding of learners' abilities across the cluster compared to their own learners (65%).

- There was no evidence of association between an assessment tool being provided and an increased belief in teachers that they were more aware of their learners' language proficiency. This may be due to the low number of respondents reporting the use of assessment tools and/or the low levels of perceived efficacy of the tools provided.
- The type of summative assessment tools provided by schools varied significantly, but over 60% of assessments were of written work, despite the consortia emphasis on oracy skills.
- Where assessment tools did exist, some teachers reported that they were created as a box ticking exercise and were not actually used in teaching practice.

2.2 Current formative assessment practices

- Despite few teachers having any formative assessment tools provided, around half (55%) did report employing formative Assessment for Learning (AfL) strategies in their Welsh language lessons (e.g. think-pair-share, facilitated peer feedback, questioning for learning).
- In interviews, teachers indicated their AfL strategies tended to be geared towards immediate assessment of and for learning within the lesson, rather than the assessment of long-term progression.
- Despite the relatively widespread use of these strategies, only 40% of teachers reported having a good or very good understanding of their learners' Welsh language proficiency.
- Teachers with higher Welsh proficiency were more likely to report having a good understanding of their learners' Welsh language competence, potentially indicating a greater efficacy in their formative assessment practice.
- Teachers who reported using formative assessment strategies were far more likely to report having a good understanding of their learners' proficiency.
- Of the teachers not using any formative assessment of Welsh (45% of respondents), 69% cited a lack of resources, 37% to a lack of Welsh language skills, and 37% to a lack of time, as the reason for not carrying out formative assessment. Only 11% ascribed their lack of formative assessment to feeling that such assessment was unnecessary.

2.4 Welsh language and transition

- Only 1 teacher reported that Welsh language assessment data was used as part of the school's transition plan. Many teachers reported that, because of this, all learners effectively go back to the very start of the Welsh Language Continuum (WLC) in year 7. Many felt this had negative impacts on learner progression and motivation.
- Some teachers felt internal transitions (i.e. between years or key stages in primary school) were handled well, with detailed handovers and efforts made to ensure learning continuity within the school.
- Many teachers raised the lack of communication, coordination and cooperation with cluster secondary schools as a challenge to creating long-term learning continuity for pupils.
- In interviews teachers often reported frustration with seeing the progress children had made in primary school ignored when they transitioned to secondary school.

2.4 Teacher attitudes to Welsh language assessment

- o Many teachers discussed Welsh being a low status within their schools, with some citing school leadership and Estyn inspections as influential in the way in which Welsh was prioritised within their schools.
- o Many teachers highlighted colleagues' low levels of Welsh language proficiency/ confidence as a factor in the de-prioritisation of Welsh within their schools and an impediment to the assessment of Welsh.
- o Some teachers reported that a lack of assessment/monitoring led to low levels of staff motivation to teach Welsh, with Welsh being prioritised lower than other subject areas that were assessed.

2.5 Teacher recommendations

- o Many teachers suggested an assessment tool that could be tailored to their class that would identify strengths and weaknesses in their learners' skills and knowledge.
- o Many teachers mentioned the need for assessments to be practical and easy to use, including by those with limited Welsh language skills.
- o Some teachers suggested some kind of assessment tool that would generate comparable data, that would allow them to understand their learners' progress towards long term educational goals.
- o Some teachers highlighted their own lack of confidence in creating assessment tools. This was supported by survey responses, where zero of the participants reported receiving any training in the assessment of Welsh.

3. Study Design

3.1 Sampling

As Welsh language assessment practice in primary education is not standardised but coordinated at an institutional level, it was important to collect data from a wide range of schools. To ensure this, sampling needed to ensure that multiple teachers from each school were not included in the study as this would over-represent the assessment practices of that institution within the sample. Participants were therefore recruited through the Welsh Language Sabbatical (WLS) courses run in Cardiff, Swansea, Aberystwyth and Bangor Universities. This ensured that participants represented a unique institution, as each teacher on the course was from a single school. Further assurance was achieved through the collection of school names as part of the survey, which were then checked for duplication.

Geographical representation was also a concern: it was important that respondents represented a broad range of socio-economic and regional contexts to ensure that discrete situated practices were not erroneously assumed to be generalisable. Using the WLS courses as a sampling frame ensured that a broad range of schools from across all regions were represented in the sample.

There are also disadvantages to the sampling approach: the WLS course is an elective training scheme that requires individuals to apply to participate and seek consent from their school. This could result in four forms of sampling bias:

- Teachers who volunteer to participate in the WLS scheme are likely to be more positively disposed towards the language and have higher levels of proficiency than the general EM teaching population. This could lead to a misleadingly positive representation of institutional policies, practices, and dispositions.
- Schools that view Welsh as a low priority are likely to be disinclined to allow staff to attend the WLS course. This could lead to an under-representation of low performing (in terms of Welsh) schools in the sample, providing a false impression of the quality and extent of assessment practices.
- It is also possible that high performing schools are under-represented in the WLS participants. Schools who are achieving extremely high standards in Welsh language education or have had numerous staff members attend the WLS course in the past, may be less likely to send staff on the course, and so be under-represented in the sample. This could lead to a false impression of lower standards or availability of Welsh assessment than exists across all schools.
- Finally, as participation in the study was voluntary, it is possible that teachers on the course who came from schools with lower levels of Welsh assessment, may be less likely to participate through a sense of embarrassment, or concern that reporting poor practice would result in punitive action from their school. This was evident in the interview participants, who were far more likely to volunteer if their school had existing assessment provision in place.

The possibility of these distortions from sampling bias must be considered when analysing the data collected and inferring trends.

3.2 Eligibility criteria & sample characteristics

Eligibility for participation in the study consisted of two criteria:

- Participants must be currently employed as a teacher in a mainstream English medium state primary or all-age school in Wales.
- Participants must have been teaching in their school for at least one academic year. This was to ensure that teachers would be aware of the institutions assessment practices, which may be specific to a particular time of year.

The data gathered on the characteristics of the teachers sampled included their self-assessed language proficiency, using the Common European Framework of Reference (CEFR) levels, and their school type in relation to the Welsh Government categories (WG 2021). Although all but one respondent represented a Category 1 (English medium) 3-11 primary school, their Welsh language proficiency was far more diverse (Fig 01), ranging from beginner (A0/1) to semi-fluent (B2):
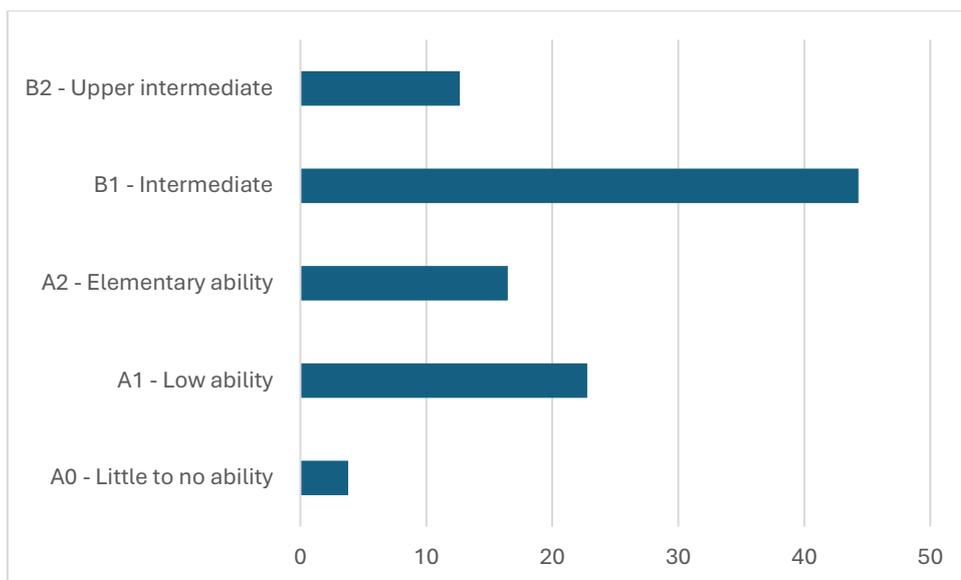
Fig 01: Self assessed CEFR levels of participating teachers (percentage values)

Although no respondent self-assessed as being in the C levels (generally considered a functionally 'fluent' language users), over half considered themselves of B1 or B2 level. Being selected from the Welsh language sabbatical (WLS) course, participants generally reflected a higher level of Welsh proficiency than would be expected in the English medium sector. For pragmatic reasons, surveys were carried out at the start of some respondents' sabbatical experience and the end of others. Many of the teachers who assessed themselves in the lower (A0/1/2) levels were likely to be at the start of their sabbatical course.

3.3 Methods

The study employed three data collection methods: teacher surveys, to provide broad quantitative data that could be subjected to statistical analysis; interviews, to provide more detailed and nuanced understanding of teachers' experiences and perspectives; and a review of assessment documentation, to provide further insights into the form and function of existing provision. The data collected were subjected to statistical and thematic analysis before being integrated through a data triangulation analysis (Flick 2011), aiming to contrast and cross-validate data from each source to achieve more robust findings (Moran-Ellis et al. 2006).

Most survey data were collected digitally through Qualtrics, with paper versions being provided to a minority of participants who did not have access to a suitable device. 81 survey responses were gathered, representing the practices of 81 different schools (approximately 9% of the number of English medium primary schools in Wales).

Consideration was given to the possibility that the wording of survey questions may bias or distort the data (Fowler 2009). Attention was paid to ensuring that the wording of the survey questions was as value neutral as possible. The survey was then piloted with five teachers not involved in the WLS scheme to ensure question wording was suitable and comprehensible, and that the interpretation of question meaning was consistent (Greig et al. 2013).

Interviews were conducted online through Microsoft Teams, or in-person on university premises. Audio files were recorded and transcripts made of all interviews. All participants who took part in the survey element of the study were invited to interview, with selection randomised in cases where more than the required number volunteered.

Of the 21 interview respondents, 18 conducted their interviews in English and 3 in Welsh. The English interviews were transcribed verbatim, except for word repetition and hesitation markers, instances of which were removed for clarity. The Welsh language interviews were translated by the researcher, and both the Welsh and the English translations of quotations are included in the findings chapter to enhance transparency (Van Nes et al. 2010).

A semi-structured interview format was selected to facilitate data production and triangulation, allowing for responsive elaboration and clarification of target themes (informed by the literature), whilst enabling the identification of new themes not covered by the interview schedule (Lamont & Swidler 2014). A responsive approach (Rubin & Rubin 2011) was employed, with participants encouraged to expand on answers or reinterpret questions to ensure they conveyed what they believed to be the most salient information. This approach was selected to provide flexibility in order to prevent artificial exclusion of participant-led themes, whilst acknowledging interviewee expertise and improving validity by better reflecting participant views and priorities (Mishler 1991). This format also allowed the interviewer to probe initial responses, deepening the quality of the data, and explore interconnecting themes (Mason 2018).

However, interview methods also have potential problems associated with them. Interviewees can feel uneasy, leading to less candid responses and the propensity towards social desirability bias (Bloor & Wood 2006). The approach also leaves open the possibility of conscious or unconscious interviewer bias, where the expectation of the interviewer influences the real or perceived response of the participant (Grieg et al 2013). Whilst attempts were made to mitigate such effects (informal setting, informed anonymity, neutral responses), it is possible that some responses were influenced by these factors, and readers should be mindful of this when considering the study's findings.

Thematic analysis was then carried out on interview transcripts using NVivo 12, with initial codes identified both inductively by their prominence and frequency within the data, and deductively by correspondence to the survey data and theoretical models from the literature (Boyatzis 1998). These codes were then grouped into provisional themes with extracts allocated to specific themes. Once all interviews had been analysed through these themes, the themes were reviewed, with some refined/merged in response to the data (Braun & Clarke 2006). These themes were then drawn into the data triangulation analysis, where they were contextualised by the survey data.

3.4 Ethical considerations

The study was carried out in compliance with the ESRC's Framework for Research Ethics (2022). The BERA Ethical Guidelines for Education Research (2024) were also considered in the study design. Ethical approval was granted from the School of Social Sciences Research Ethics Committee on the 25/03/2024 (SREC reference REF).

Informed consent to participate is an essential element of research ethics (Hammersley & Traianou 2012) and numerous steps were taken to ensure that participants were fully aware of the research goals, processes and potential outcomes prior to agreeing to participate. Before participating staff, learners, and parents received detailed information regarding the study's objectives and methods. All information was made available at least a week prior to the start of data collection to ensure participants had ample time to review the information and come to a decision.

Opt-in consent was required for both elements of the study (surveys and interviews). Consent for survey participation was integrated into the survey design and recorded as part of the survey response. Interview consent required the completion of a consent form prior to the interview taking place.

Provision was made for participants to withdraw consent after data collection, however no participants requested this. In addition, steps were taken to protect the identity of participants, the security of their data, and ensure the physical and emotional wellbeing during data collection. Protecting participant identity was taken very seriously, in accordance with the Data Protection Act (UK Government 2018). Participant identity was anonymised on site, with each participant self-allocating a pseudonym, allowing the linking of different data formats whilst ensuring only anonymised data was recorded.

Interview recordings have been stored securely on the University's One Drive using participant pseudonyms only. Any names or other identifying information have been redacted from the transcripts, along with data that could identify the school.

4. Findings & Discussion

The findings of the study fell into five themes: the assessment tools provided by schools (both formative and summative); the formative assessment practices used by teachers; the way assessment was used to facilitate transition; the attitudes of teachers and others in their schools towards the assessment of Welsh; and the recommendations made by teachers to enhance the assessment of Welsh.

4.1 The provision of assessment tools and resources

The study explored the provision of assessment tools through a series of survey questions. Participants were asked whether they had access to Welsh language assessment tools, with examples of possible resources highlighted. If participants responded that tools were provided, they were asked to clarify the form that the assessment took (e.g. marking rubrics, marking criteria for reviewing written work, level guide for oracy performance). If participants reported receiving tools for the assessment of Welsh, follow up questions were used to identify how effective and practical they felt the resources were in meeting learner needs. Finally, participants were asked whether the tools provided were shared with other schools in their cluster, or if they were distinct to their school.

These questions were followed up in interviews, with participants asked to explain the use of provision, and express their opinions concerning its efficacy, accuracy and practicality. Contextual factors were also explored including school demographics, cluster interaction and coordination, prevailing attitudes amongst staff and learners, and support from school leadership for Welsh language teaching.

Participants also supplied examples of the assessment tools provided by their schools for analysis including marking criteria, rubrics, coverage lists, and learner language maps.

The main findings concerning the provision of assessment tools can be divided into five areas, prevalence, efficacy, accuracy, subjectivity and comparability. For the purpose of clarity, these are discussed below separately, though there is obviously a significant degree of interconnectivity between these themes.

4.1.1 Prevalence of assessment tools

It is important to start by contextualising the prevalence of Welsh language assessment tools in English medium primary education, before considering the characteristics of those that are in use. The study data indicates that the majority of schools currently have no tools for the assessment of Welsh.
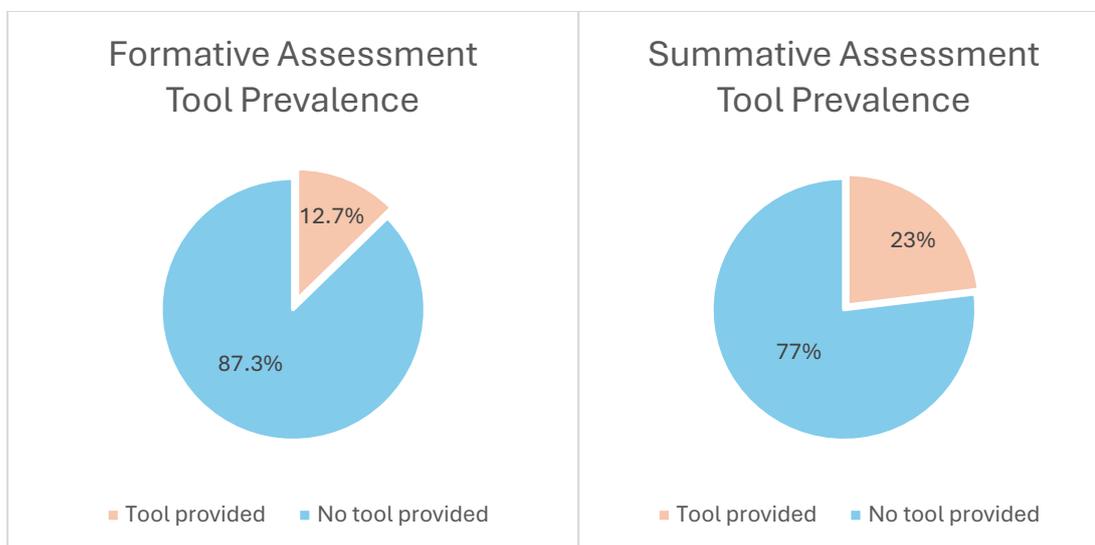

Fig. 02: Pie-charts showing the prevalence of formative and summative assessment tools

The majority of schools appear to no longer carry out summative assessment of Welsh, however the greater preponderance of summative assessment tools (compared with formative assessment tools) is likely reflective of historic statutory requirements to report Welsh outcomes at the end of Key Stages (WG 2019).

In interviews some teachers linked the lack of assessment with a de-prioritisation of Welsh within their schools:

> It's the scores, isn't it? That can be shown against other schools. I think that's why maths and English have been put first.
> *Interviewee 07*

> Welsh is seen as something that you've got to tick-off. I've done my 15 minutes.
> Yeah, I think so, unfortunately. The timetable is too full, it's bursting at the seams.
> As teachers, we're constantly asked, they want data on maths and English and that's what parents want to know about.
> *Interviewee 12*

> I think it's because there literally are assessment systems [for maths and English] and an understanding of how to assess already in place, whereas with the Welsh, I think we're still in a situation where some of the staff are still getting to grips with the Continwwm y Iaith [the content structure provided by the consortium, in this case CSC).
> *Interviewee 16*

This relationship between assessment and status is likely to be bidirectional (Bleazby 2015), with a lack of assessment causing a de-prioritisation of Welsh, whilst a low status of Welsh within a school potentially decreases the likelihood of that school carrying out Welsh assessment. The specific prioritisation of mathematics and English has been observed to have

a 'narrowing' effect of curriculum time and subject status in England (British Academy 2024; Quick 2024) and internationally (Berliner & Nichols 2007; Ewing 2010). This theme is explored in more detail in section 4.4.

A small number of staff who reported having an assessment tool provided, highlighted the issue of the deployment of assessment provision:

> I've responded that we do summative assessment, but as far as I know the assessment actually only exists on paper. We made it for an Estyn inspection and I don't think it's even been looked at since. Teachers just don't have time to do lengthy assessments, and it's never monitored.
> *Survey response from 'Betty'*

> When I knew I was going to chat to you today I checked and turns out that we do have an assessment for Welsh, but I've been at the school, you know, for 3 years and this was the first time I'd seen it.
> *Interviewee 19*

This element of deployment was not specifically included in the survey questions, only being raised in elective responses, so it is not possible to infer the prevalence of such superficial engagement with Welsh language assessment. Further research into this would indicate whether this reported issue is widespread, and whether it is an unintended result of previous statutory requirements (WG 2019) or inspection expectations.

4.1.2 Efficacy of formative assessment tools

Before exploring the data around formative assessment tools, it is important to highlight that this data is distinct from formative assessment practice (as explored in section 4.2). Formative assessment is often embedded in teacher pedagogy and does not necessarily require external tools. For example, formative assessment strategies may include the use of questioning to track progression and encourage engagement, facilitating peer assessment, or monitoring task outputs. However, formal formative assessment tools have been found to be beneficial in allowing teachers to track longer term learning trajectories and direct remediation (Connor et al. 2007; Kowalski et al. 2015).

Teachers who reported being provided with a formative assessment tool by their school (approx. 13% of participants), were asked to reflect on the efficacy and practicality of the provision. This analysis was prompted in the survey through an itemised qualitative matrix of formative assessment purposes (Black & William 2009): informing differentiation, directing use/design of scaffolding, targeting remediation, influencing pedagogy, selection of provision, and directing the pace of progression.

Overall teachers' opinions around the efficacy of the formative assessment tools provided in meeting the needs of learners were largely negative (Fig. 03), with the majority believing that the tools provided performed less than 'moderately well' and no teachers believing that the tools provided performed 'very' or 'extremely well'.
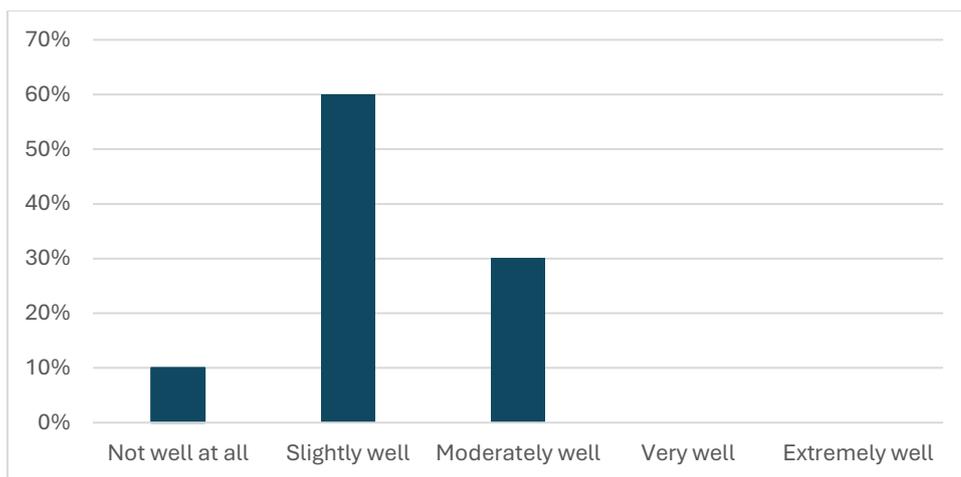
Fig. 03: Teacher feedback on the performance of assessment tools provided.

In interviews, the issue of efficacy was discussed with teachers who tended to express polarised perceptions of the formative assessment provision provided. Some felt that existing provision met the needs of the learners, whilst others reported that the tools failed to provide useful data for the directions/adaption of teaching and provision. The quality and purpose of the feedback was raised by some respondents, who highlighted a lack of specificity that would otherwise facilitate effective targeting of remediation, or inform differentiation, and a misdirection of feedback towards leadership and inspection bodies rather than learners and teachers:

> The feedback we usually get is like, 'Oh, that was good', or 'They were very confident', or 'They did well', or 'They answered a lot of questions'.
> *Interviewee 18*

> [The assessment data] is not for the actual people that have been doing the learning. But for people who are who are interested or need to see something. So, you know Estyn, other practitioners, parents on, you know, on parents' evenings. I don't think, it's not primarily for the actual children who have done that work.
> *Interviewee 20*

This perceived need to evidence for education stakeholders was also cited with reference to the emphasis on written tasks in some schools:

> IR: What was the impetus for that [putting written work in books]
> IE: Well, just trying to get evidence in the book for Estyn, isn't it. There was a lot of pressure from the top to get, you know, 'There's no Welsh in your books'.
> Interviewee 02

> IR: Why do you finish up with a written piece rather than, say, a spoken piece or something like that?
> IE: I think it, it comes from a sort of evidence point of view... it's a tangible piece of evidence as to what has happened in the learning journey in the in the language journey.
> Interviewee 20

Such external motivations for carrying out assessment were often seen as impositions that did not focus on facilitating learning, but instead added to teacher workload (see section 4.4)

4.1.3 Accuracy of summative assessment tools

Summative assessment of Welsh included a broad range of approaches (Fig. 04). Such plurality is expected given the CfW emphasis on targeted and bespoke assessment designed to meet the needs of individual learners and cohorts (WG 2024). However, the prevalence of assessment based on written work does not align with the emphasis regional consortia have placed on the development of oracy skills at primary school age.
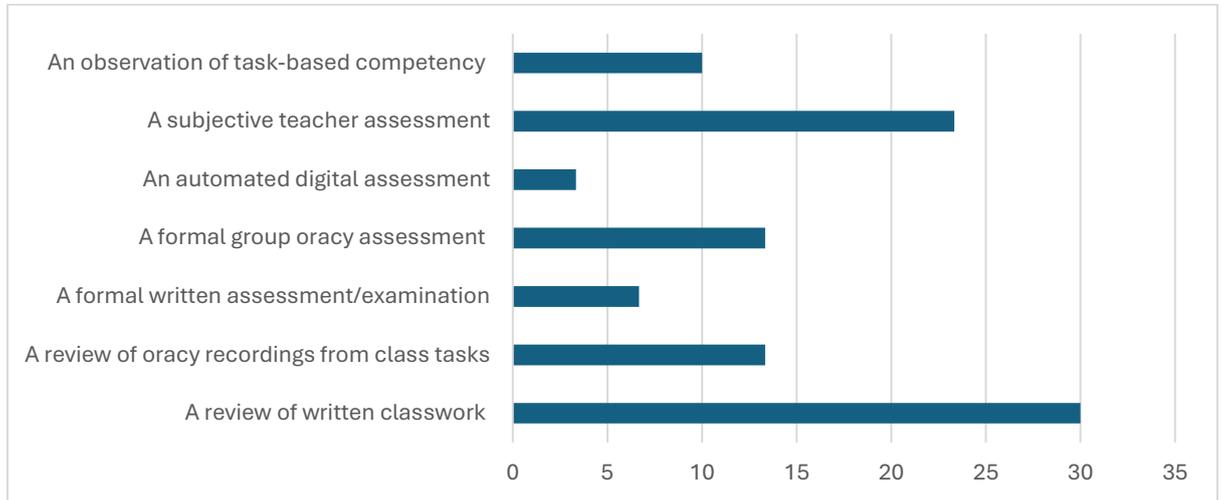
Fig. 04: Percentage of summative assessment types reported in survey responses. Respondents were able to select multiple options and submit their own responses.

Teachers who reported being provided with a summative assessment tool by their school were asked to reflect on the accuracy of the provision in representing learner competence. In the survey data this was achieved through a unidimensional scale ranging from 'not accurate at all' to 'extremely accurate'. The distribution of responses indicated a generally low level of perceived accuracy in the assessment's ability to reflect learner competency (Fig. 05).
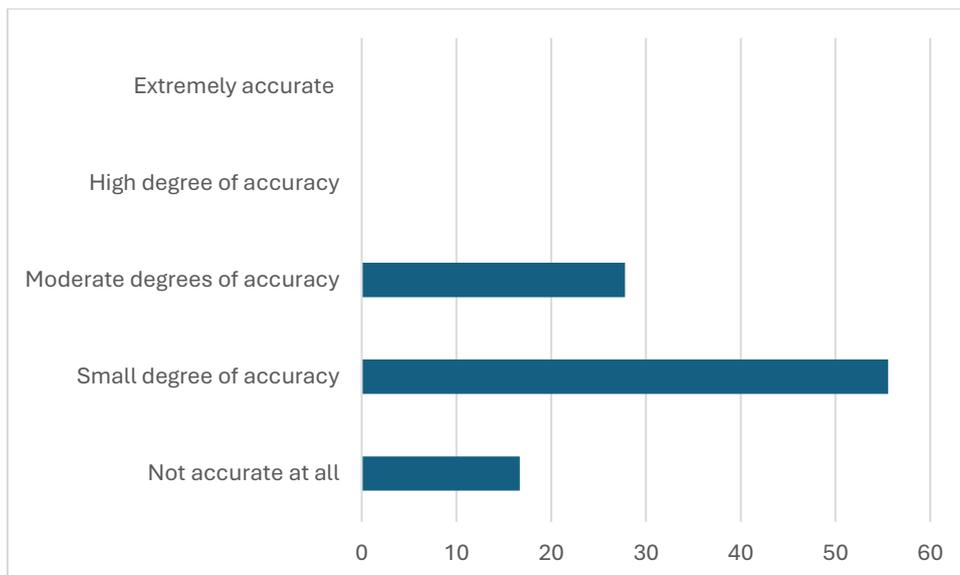
Fig. 05: Teachers' perceptions of the accuracy of the summative assessment provided by their schools (percentage values)

This theme was explored in teacher interviews with three key factors being identified: the level of detail/scope, the targeted domain of language competency, and the authenticity of the tasks.

In terms of the level of detail, respondents reported that summative assessments tended to be either very superficial with teachers asked to grade learners' ability without reference to any objective scale or measure (39% of survey respondents), or very high detail but low scope, with learners questioned on every pattern in the school's curriculum, but with only a small number of selected students assessed:

> So, they [the teachers] are going to choose the three highest ones [pupils]. So, that's not very reflective of the class.
> *Interviewee 07*
>
> So, it's usually 6 [pupils from each class]. Me personally, I always ask which children are confident in Welsh, and then I kind of send children off depending on who feel competent to speak... I know within the school that there are some teachers who prefer to send their more able.
> *Interviewee 18*

Either approach was felt by respondents to compromise the accuracy of the assessment scores and not accurately reflect the range of abilities within a given class.

Domain factors tended to be raised where teachers felt that summative assessment focused too heavily on written work, therefore not accurately reflecting long-term knowledge retention (e.g. where written assessment included heavy scaffolding), or more holistic language competence (i.e. inclusive of factors such as phonological competency, generativity, and automaticity):

> There's no point, right? Because they [the pupils] were writing in their books, they were showing me the books with all this lovely writing in Welsh and I was like, 'Well, how much do they actually know what it means?'
> *Interviewee 17*

Concerns about assessment authenticity were raised with regards to the potential misrepresentation of learner competence. Some teachers felt existing assessments misrepresented learner competence in the way they allowed specific scaffolding and preparation for the assessment task (e.g. memorising a script by heart, or preparing a prompt-sheet), therefore failing to be representational of learners' ability to use the language spontaneously or communicatively.

> They are given things like language mats, and they've got access to Welsh dictionaries... It doesn't take into consideration their oracy, their reading, and their Welsh off-the-cuff kind of thing... it's a good way to showcase what they've learned, but I do think it doesn't cover everything with Welsh language.
> *Interviewee 20*

In analysing the assessment tools provided by participants, it became clear that some teachers fail to make a distinction between assessments of proficiency and the tracking of coverage. Some of the examples provided consisted of a list of language patterns that the teacher would tick off once covered in a lesson. This obviously does not reflect learner competency, and the perception that this type of coverage tracker fulfils the assessment needs of learners may lead to a failure to develop any actual assessment processes.

This conflation of coverage and proficiency was a more general theme in the existing provision, with assessment tools often covering the school's entire Welsh language curricular content. This means that scores were a compound of learner proficiency and content coverage. This lack of discretion makes it difficult for teachers to identify whether learning deficits are the result of misaligned coverage, or failures in learning/retention of content (Glaser et al. 2001).

Summative assessment accuracy obviously has significant implications for the efficacy of inter and intra-school transitions (discussed in section 4.3), with misrepresentation of learner proficiency compromising teachers' ability to provide the learning continuity that is essential for language acquisition (Fletcher 2018; Bolster 2009; Braund 2009). The generally low levels of prevalence and perceived accuracy raise concerns about the extent to which such continuity is currently being achieved, potentially leading to high rates of learning attrition (Russell 2024).

4.1.4 Subjectivity of assessment tools

Another theme evident in the analysis of current assessment tools was a significant reliance on subjective assessment. This was evident in examples of both teacher and self-assessment tools, which often required a qualitative judgement around confidence or competence in a particular skill or pattern use. Of the teachers who reported being provided with assessment tools, around 40% specified subjective assessments as contributing to their school's assessment protocol.

Whilst such formats have advantages in terms of practicality and time efficiency (Andrade & Cizek 2010), and are capable of enhancing learning (Andrade & Valtcheva 2009), they are susceptible to a variety of biases and distortions, including:

- Overestimation bias – where learners tend to overestimate their abilities in self-assessment formats. This effect is not generally evenly distributed, with some MAT learners susceptible to 'modesty bias' leading them to underestimate their performance (MacIntyre et al. 1997). Over-estimation bias can also apply to subjective teacher assessment, with staff likely to overestimate the ability of their own students (Schmidt et al. 2023). Students' behaviour, personality, appearance, ethnicity, or perceived motivation can all impact on staff assessments of competence (Cumming 2001; Rea-Dickins 2001). This often manifests as a 'halo effect', where overall impressions of a student influence assessment ratings (McNamara 1996). This inconsistency makes it challenging to accommodate overestimation bias in assessment.
- Dunning-Kruger effect – the tendency of learners with lower ability to disproportionately overestimate their performance in self-assessment due to a lack of awareness of their own mistakes or inadequacies (Kruger & Dunning 1999; Gaffney 2018)
- Social desirability bias – where learners may deliberately overstate their abilities because they believe it is what their teachers want, or in response to perceived institutional expectations (Boud & Falchikov 1989)
- Self-promotion bias – where the learner deliberately inflates their self-assessment score as a strategic form of self-promotion (Leon et al. 2021).
- Demographic bias – studies have shown that factors such as cultural background (Heine & Hamamura 2007) and gender (Liu 2018) may influence self-assessments of proficiency.

While self-assessment has an important place as part of a holistic suite of assessment (Black & William 1998), in many schools this was the only form of assessment used. Concerns about this

kind of subjectivity in assessment practices were highlighted by teachers in the interviews, who recognised both the inconsistency and the situated nature of such assessments:

> A lot of those sort of formal summative assessments, become box ticking and so on, because it is down to the teachers' interpretation of what they can do, and in a in an ideal world it wouldn't be like that.
> *Interviewee 14*

> Sometimes you were holding your finger in the air and thinking, well, yeah, I think, I guess they can do that on a good day.
> *Interviewee 01*

> I think this comes down to individual teachers who could see it in a different way. They can interpret what a child can do in a different way.
> *Interviewee 13*

Subjective teacher assessments are inevitably influenced by personal experience and institutional factors that lead to context-dependent scoring, where learners are scored on the basis of their proficiency in relation to their peers rather than an objective measure. This reflects Antibi's (2003) 'macabre constant', which posits that in subjective assessment teachers unconsciously allocate a fixed proportion of grades across a performance spectrum, potentially leading to systematic over/underestimation of student performance. Whilst adequate for internal levelling (e.g. for streaming), and a legitimate tool for formative assessment, such a lack of objectivity leaves performance scores unsuitable for inter-institution comparison or objective measurement of performance.

4.1.5 Comparability of assessments

As already touched upon in the sections on accuracy and subjectivity, a major theme in the assessments currently used in English medium schools is a lack of comparability. The comparability of assessment data operates at two levels:

- Intra-comparability, i.e. the comparability of longitudinal data formed by a series of scores for an individual
- Inter-comparability, i.e. cross-sectional comparability of scores between individuals/cohorts/institutions (Robinson et al. 2005).

Many of the school assessments examined showed relatively high levels of intra-comparability. Examples often took the form of a standardised linear assessment that encompassed the whole school curriculum content, allowing the same tool to be used across year groups and provide comparable data on learning progression. This is reflective of the CfW emphasis on assessments as a facilitator of individual progression (WG 2024). Such an assessment model does however risk damage to learner self-efficacy through poor alignment with content coverage (Biggs 1996). For example, in some cases assessments are aligned with the broad progression steps, rather than year group, resulting in content that fails to align with teaching and learning.

However, none of the assessments reviewed demonstrated high levels of inter-comparability, due to variations in content, format, difficulty, scoring, and target constructs. In order to be comparable, the assessment design must ensure equivalence of constructs, establish validity, and maintain reliability (Bennett 2011). This is a demanding set of requirements for teachers and schools who are largely required to develop their own assessments without additional

specialist training. Whilst the delegation of assessment to schools is intended to be positive, allowing schools the autonomy to develop tools that fit their context and needs (WG 2024), some teachers voiced frustration and a desire for support in the assessment of Welsh:

> [We don't have] anything. There is nothing used across the school as far as I know.
> *Survey response from 'Sam Tan'*
>
> I feel, the new curriculum, it's always been assessment that has failed it, you know? We've never been given anything to go on, other than these 'what matters statements'.
> *Interviewee 02*
>
> I've had no training on the development of assessments, let along specifically the Welsh language assessment.
> *Interviewee 19*

A simple solution to the problem of inter-comparability is for schools to adopt the same assessment across clusters. Whilst assessments were shared amongst other cluster schools in a small proportion of cases (6% of summative assessments), for the most part, tools were developed internally, meaning the assessment was unique to that institution. Therefore, even in the minority of cases where assessments existed, teachers had very little awareness of how well their learners were performing in relation to learners from other schools (Fig. 6), with only 3% of respondents believing they had a good understanding.
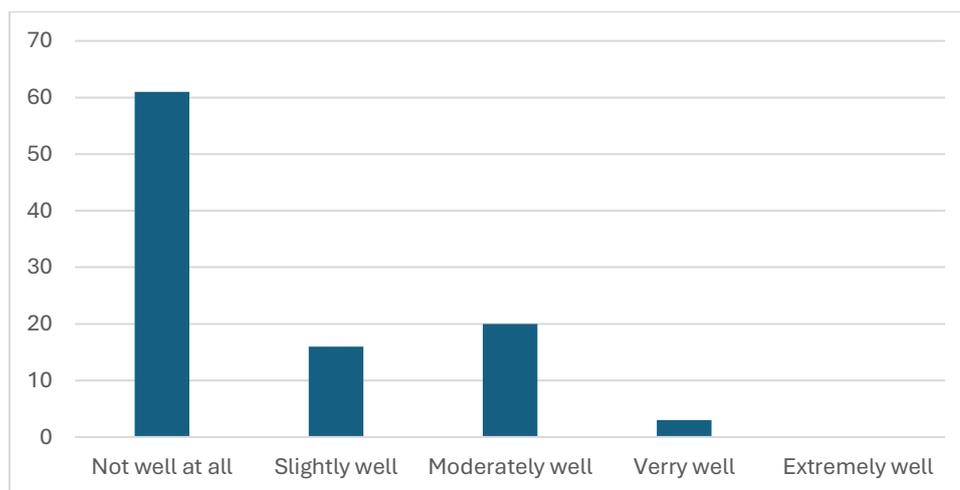
Fig. 6: How well teachers felt they understood the proficiency of their learners in relation to learners in other schools (percentage values)

This was supported in interviews, with many teachers discussing this very situated understanding of learner proficiency:

> We used to have moderation and standardisation with other schools, which is really, really beneficial because, you know, since COVID you can be stuck in your own little bubble and you could think, 'Oh, yeah, we're amazing at this', or, 'We need to progress at that'. But unless you've got another benchmark and a professional dialogue with other schools, it's really impossible to tell.
> *Interviewee 02*

IR: Os ti'n meddwl o safon o'r plant yn dy ddosbarth, wyt ti'n cael synnwyr o pa safon ydyn nhw, o'i gymharu a safon o plant o ysgolion eraill?
IE: Dim o gwbl, ni'n byw mewn bubble.

IR: If you think of the level of the children in your class, do you have a sense of what level they have compared to the level of children from other schools?
IE: Not at all, we live in a bubble.
Interviewee 11

However, collaborative assessment practice was evident in a small number of schools, with teachers reporting a moderation process within their cluster to try and create consistency and comparability in levelling Welsh learners:

We meet between three and five times a year as a cluster. We take learners of different levels in their learning and they're compared across the board. So, we may be comparing oracy videos, written Welsh, we compare books against everyone else.
Interviewee 04

Whilst moderation has the potential to greatly improve the inter-comparability accuracy, it demands a significant allocation of time and staff resources (Miller 2025), raising practical and financial challenges for schools. Such cluster-based moderation could also lead to an institution-level manifestation of the 'big fish little pond' effect (Marsh 1987), thereby driving standards down rather than up. Some teachers reported a drop-off in cluster-based coordination since the Covid pandemic and in response to the implementation of the CfW:

Yn anffodus, roedden ni'n arfer cwrdd am cyfarfod 'moderation' gyda ysgolion eraill, ond dwy e ddim yn digwydd nawr dan y cwricwlwm newydd. Mae anodd i wybod beth mae ysgolion eraill yn wneud.

Unfortunately, we used to meet with other schools for moderation meetings, but it doesn't happen now under the new curriculum. It's difficult to know what other schools are doing.
Interviewee 08

This lack of collaboration in Welsh assessment seems to be mirrored in the school's relationships to their cluster secondary: only one respondent reported sharing Welsh language assessment information with their associated secondary. This has implications for transitions (discussed in section 4.3) but is also a factor in terms of comparability. Secondary schools providing information on expectations of learner knowledge and performance could be highly beneficial in helping cluster primaries to develop a shared understanding of standards and aspirations of learner proficiency.

4.1.6 Conclusions concerning the provision of assessment tools

The data indicates that the majority of participating teachers believe there are currently significant deficits in the provision of appropriate and effective Welsh language assessment tools in the English medium primary sector. The majority of schools have no Welsh language assessment provision available for teachers. In cases where schools have provided assessment tools, the provision is often considered inappropriate or inadequate by teaching staff, with existing assessments sometimes not being used at all. When they are used, the impact of tools provided are often impaired by deficiencies in accuracy, consistency, and comparability, making them poor measures of inter-comparability and limiting their utility in providing actionable data.

There are however examples of good practice, with school clusters and consortia collaborating to produce effective assessment tools, but at present these instances appear to be the exception.

4.2 Formative assessment practices

As opposed to the more formal formative assessment tools explore in section 4.1, formative assessment practice refers to an ongoing process of evaluating student learning for the purpose of informing teaching and supporting future learning, rather than merely measuring summative outcomes (Bachman and Palmer 1996). It involves teachers and students continually gathering evidence of proficiency and using this information to adjust instruction and address students' learning needs, using assessment as a means of facilitating and enhancing learning (Black and Wiliam 1998). This approach allows teachers to identify gaps in knowledge or skills early, fosters self-regulated learning through feedback and reflection, and has been shown to improve student achievement (Foster 2024). Formative assessment is typically facilitated in classrooms through techniques such as diagnostic questioning, feedback dialogues, peer and self-assessment activities, and quizzes (Carless 2005; Wiliam 2011).

4.2.1 The prevalence of formative assessment

As expected, formative assessment strategies were more widely reported than the use of specific assessment tools. Formative assessment is a foundational pedagogical approach that is used by most teachers across the whole curriculum. However, although more prevalent than the use of formal assessment tools (55% compared to 13%), the number of teachers using formative assessment strategies in the delivery of Welsh provision is significantly lower than one might expect compared to other subject areas.

In the survey data, teachers identified a number of reasons for not using formative assessment in their Welsh teaching (Fig. 7):
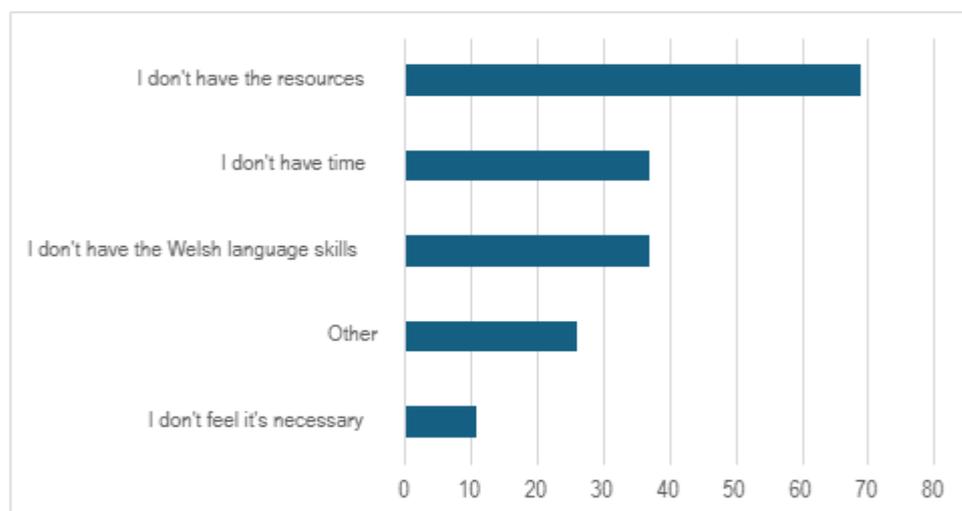
Fig. 7: Reason for not using formative assessment in the delivery of Welsh (values show the percentage selected - participants were able to select more than one option and submit their own responses).

The most common response was that they did not feel they had the resources to carry out formative assessment (69%). This is surprising as most formative assessment does not require the use of any specialist resources. It is likely therefore that the selection of this item stems from a misapprehension amongst respondents about what constitutes formative assessment, raising the possibility that formative assessment approaches are more common than the initial

response suggested. It may also be reflective of the previously mentioned challenges posed by teacher Welsh language proficiency, resulting in teachers feeling that their lower confidence in Welsh necessitates the deployment of assessment resources.

Respondents also selected other factors that were influential in the limitation of formative assessment in their practice. These included time pressure and a lack of Welsh language proficiency. In open responses, respondents also highlighted a high proportion of EAL learners in the classroom, a lack of assessment training, and the responsibility for Welsh lying with the PPA (cover) teacher, as reasons for not carrying out any formative assessment of Welsh. Only 11% of respondents not carrying out formative assessment attributed this to a perception of formative assessment as unnecessary.

Although there was no association in the data between the participants Welsh language proficiency and the likelihood of reporting the use of formative assessment, a lack of skills and training in colleagues was a common theme raised in interviews. Whilst participants tended to have comparably high levels of both language and pedagogical competence in Welsh, they reported that these factors were a significant challenge for other staff members:

> Teachers don't find it as easy to do [formative assessment] in Welsh, sort of identify the next steps and show, you know, model that to the children for them to be able to do that themselves.
> *Interviewee 06*
>
> IR: Would you feel as confident then doing formative assessment in a Welsh lesson? To understand your learner's abilities?
> IE: Now, yes. After this course, 100%. Prior to it, I wouldn't even necessarily differentiate the Welsh lesson. It's just kind of we all muck into it. But that's through not being a Welsh speaker.
> *Interviewee 12*
>
> When you have a lack of confidence in something, especially language, immediately you're less comfortable in using familiar [AFL] strategies. Because I know from before when I was a lot less confident when speaking Welsh, I would feel I genuinely nervous assessing a child, when I am not 100% sure that I'm doing it right. Because I don't want to assess someone and tell them, give them feedback, but the feedback is not entirely accurate. So, it definitely effects other staff.
> *Interviewee 20*

This was supported by the survey data, which showed a significant association (p=0.019) between perceived teacher language proficiency and familiarity with the ability of their learners (appendix item 1). Teachers who self-assessed as having a CEFR level of B1 or higher had 65% greater odds of believing they knew their learners' Welsh language proficiency moderately well or better, compared with those at level A2 or lower. So, whilst the use of formative assessment does not appear to be linked with participant proficiency, the efficacy of that assessment in appraising learning is significantly associated.

Schildkamp et al. (2020) emphasise that deep content knowledge is required for teachers to interpret student responses accurately and provide meaningful feedback, a fundamental aspect of formative assessent. In a language context this can manifest in the efficacy of modelling, correction, explicit explanation, and monitoring of progression (Ellis et al 2006; Li & Zhang 2016; Graham et al 2017). A deficit in Welsh language proficiency is therefore likely to

compromise a teacher's ability to deliver optimal formative assessment. However, Yusuf and Novita (2020) do find that although teacher language proficiency impacted upon lesson content, it did not compromise teachers' ability to assess and provide feedback to learners at lower levels of language development. Unfortunately, the survey data did not include in which progression step each participant was working, preventing the exploration of the potential stratified impact of language proficiency on formative assessment.

The Welsh language proficiency deficit in teachers was also felt by some respondents to be exacerbated by the delegation of Welsh to a single colleague:

> They give it [Welsh] to the PPA [cover teacher]... Well actually, you don't know then what's going on in your class for. You should know what Welsh looks like in your class, not be asking me.
> *Interviewee 17*
>
> I think it [the PPA teacher taking responsibility for Welsh] is quite detrimental really... I think as much as other teachers would view it as easier for the Welsh leader to take over, I think in regards to the plan of the future, I think it needs to change... it sounds awful. It's not like that, but it's all sort of on the Welsh lead really and unless in specific information is needed, the rest of the staff will just be doing drill sessions each day and that is it. There's no set plan for that.
> *Interviewee 20*
>
> I predominantly teach the Welsh and so then the other staff don't think they have to do it. So, I go in once a week and teach Welsh. I do it an hour, hour and a half if I'm lucky... and the other staff think, 'Oh well, she's doing that. I don't need to do it'.
> *Interviewee 21*

This can result in teachers removing Welsh from their individual practice, losing a sense of their class's progression in the language and failing to embed the language in pedagogy as advocated by the consortia. Whilst only reported in a minority of cases, the implications of such practice are significant, and should be considered a cause for concern in any appraisal of a school's faculty for deploying Welsh learning across their curriculum.

4.2.2 Types of formative assessment

Participants who reported using formative assessment in the survey were asked to specify which kinds of assessment they used. The responses indicated a wide range of approaches covering all the most common forms of formative assessment (Fig. 8):
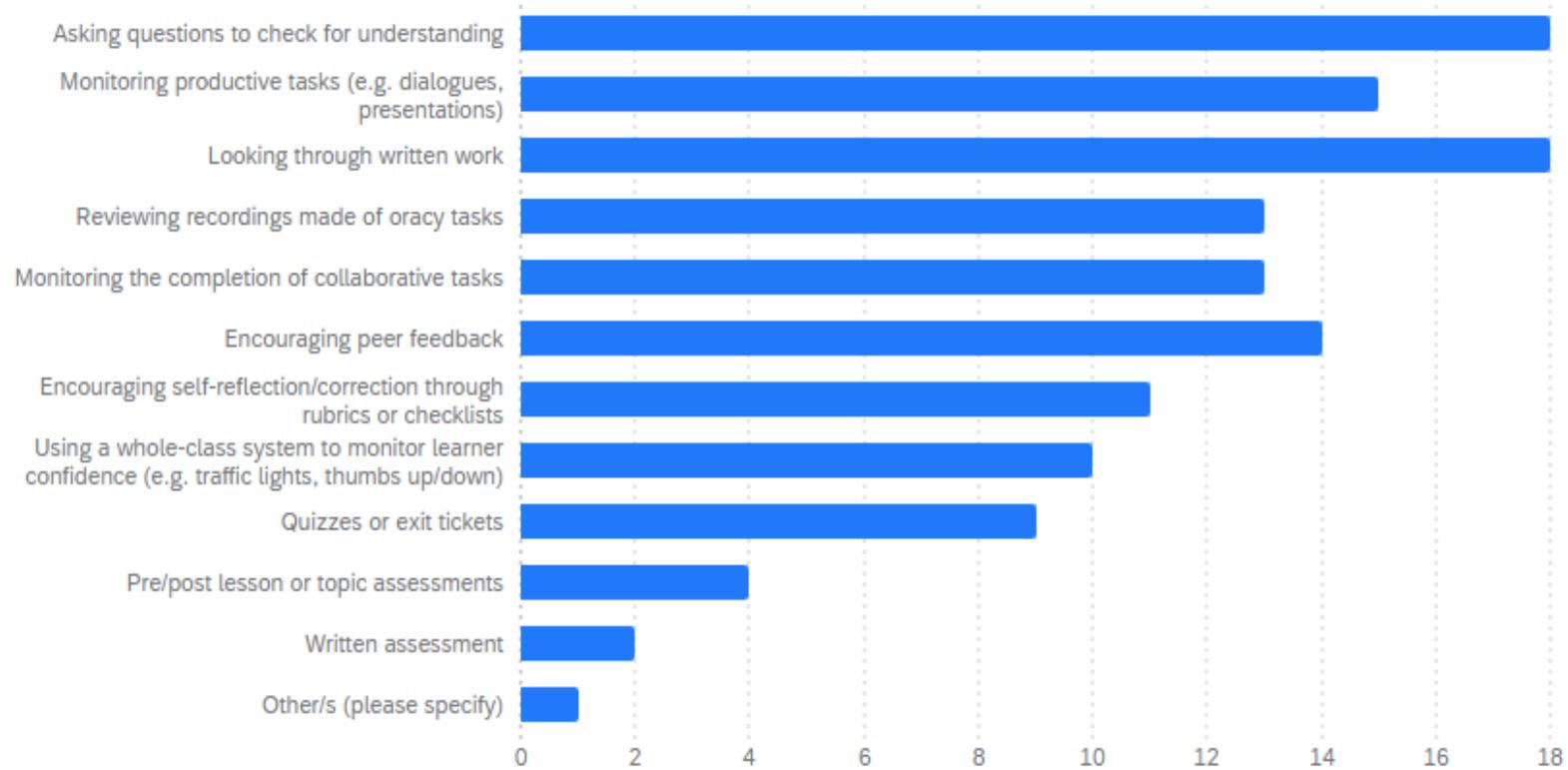
Fig. 8: The kinds of formative assessment carried out by respondents (participants were able to select more than one option)

One notable feature was the relatively low number of respondents who reported carrying out pre/post topic assessments. This is potentially indicative of an emphasis on immediate assessment of learning. Whilst timeliness is a central feature of formative assessment, it can potentially lead to a failure to assess long-term learning (Bennet 2011; Coffey et al 2011). This is particularly pertinent to the early stages of second language teaching, where the development and curation of a foundational lexicon is essential in combating attrition and developing the lexical-syntactic knowledge required for continued progression (Nation 2001; Webb & Nation 2017).

Despite this potential omission, respondents using formative assessment is likely to contribute significantly to their understanding of their pupils' Welsh language competence. This is supported by a regression analysis of the survey data, which demonstrates that teachers who report using formative assessment are more than three times as likely to feel they know their learners' ability moderately well or better (p=0.023) when controlling for self-assessed Welsh language proficiency (appendix item 2).

4.2.3 Impact of formative assessment

Survey participants were asked to rate the impact of their formative assessment practices on their pedagogy, lesson content and learning provision. Whilst many respondents felt that their formative assessment did impact significantly on their pedagogy (i.e. differentiation, pacing, remediation), this was less true of lesson content and learning provision (Fig. 9):
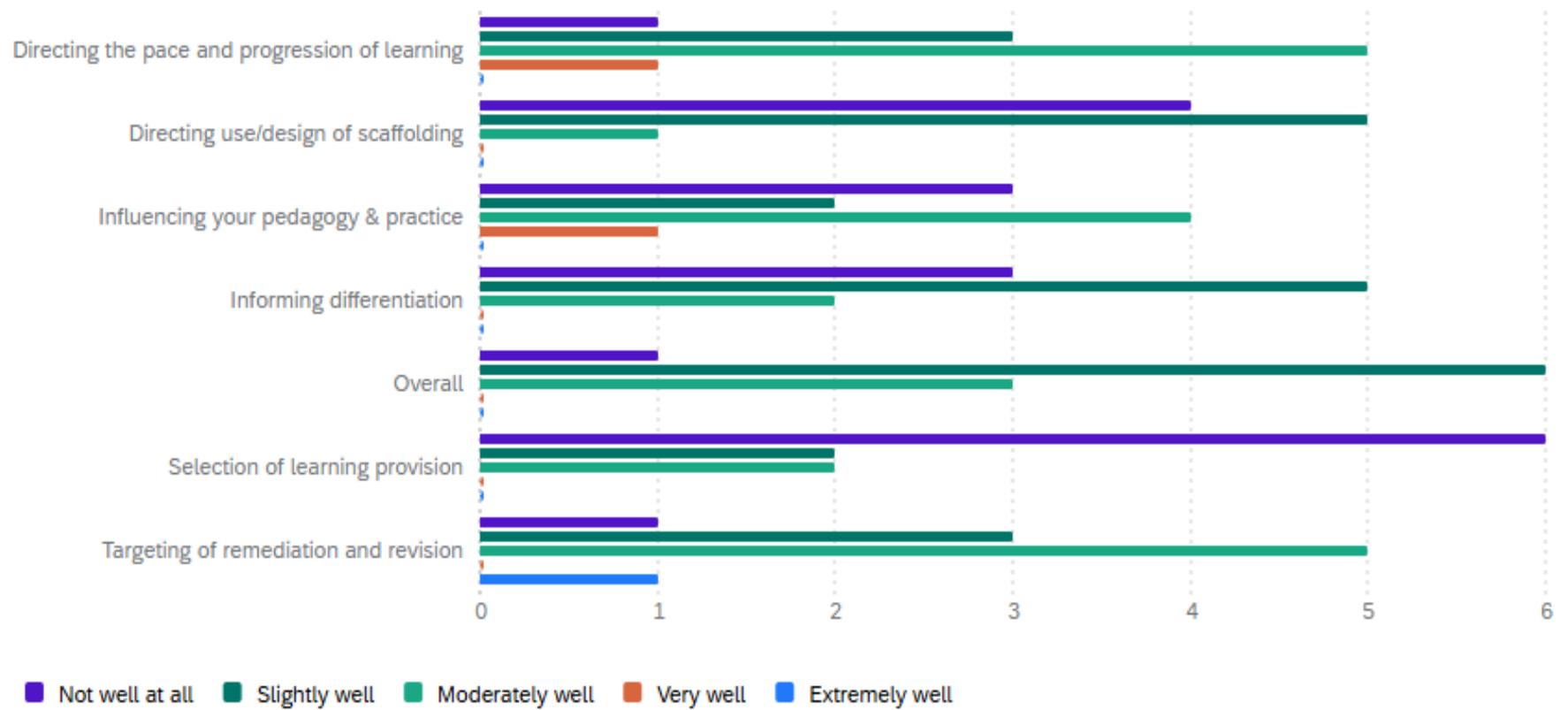
Fig. 9: How well teachers felt their formative assessment met the needs of learners by area.

As pedagogy is something that can be adapted immediately in response to learning progression, whilst adapting provision and lesson content often requires the carrying over of assessment data from one lesson to the next, this could be indicative of the more short-term focus discussed previously in section 4.2.2. This short-termism in formative assessment approaches was discussed by some teachers during interviews:

> It just isn't logged or recorded or shared in any way, if that makes sense. So, I think the teachers are able to differentiate and I think they are able to keep a track whilst they're teaching... You know, it's only happening sort of in their short-term planning and it's only impacting their short-term planning, but they're probably not then considering the end goal.
> *Interviewee 06*

> I think there's definitely AFL strategies being used in lessons, but for me that's kind of a habit, you know, I asked the questions, think pair share, all that, I do it in all my lessons, but does it actually change what I teach? I dunno, I think I mostly just like go with the lesson plans on the drive [the school's shared computer drive].
> *Interviewee 19*

Such limited or habitual use of formative assessment is highlighted by William (2018), who draws a clear distinction between what he describes as 'formative intention' and 'formative action': using the tools of formative assessment does not necessarily imply that a teacher allows them to impact upon their practice. Teachers can, in effect, 'go through the motions' of formative assessment, whilst allowing it to only superficially affect their pedagogy, if at all. This is a challenging phenomenon to explore as such 'con-formative' behaviour (as opposed to formative action) is often unconscious (Heritage 2007; Torrance 2012).

The effect of a lack of recording of ongoing assessment could be easily overlooked. Formative assessment is primarily 'live' in the sense of being immediate and responsive to learning and learner needs (Black & William 1998). However, this does not lend itself to formative action over a series of lessons and over longer time periods. To affect such extended scope, formative assessment benefits from the integration of some form of record of learning progression in order to inform the planning of future input (Carney et al. 2022). However as explored in section 4.1, these are currently only available in a small number of schools.

4.3 Assessment and learner transition

Assessment plays an important role in supporting learners through key transition points (e.g. primary to secondary education, or between key-stages) by ensuring continuity of learning and informing appropriate pedagogical responses. Accurate assessments of learner competency help to bridge gaps between educational stages by providing teachers with a clear understanding of pupils' prior attainment, learning needs, and potential barriers to progression (Galton et al. 2003). Research indicates that poorly managed transitions, compounded by a lack of assessment information, can lead to a decline in academic motivation and achievement (Jindal-Snape & Foggie 2008; Evangelou et al. 2008).

Welsh Government policy requires schools to agree and implement a Transition Plan that ensures the continuity of learning for pupils (WG 2022, Regulation 6(3).2). This element of learning continuity is also emphasised in the CfW guidance on assessment (WG 2024b):

> Effective transition is about supporting all learners along the learning continuum, as they move between different groups, different classes, different years and different settings. Ensuring the well-being of all learners should be an important and integral part of the process, recognising the needs of individuals, while also supporting both continuity and progression in their learning.
> *WG 2024*

This focus is pertinent in second language acquisition generally, where learning continuity is of particular importance (Pfenninger & Lendl 2017), with transition points posing a particular challenge (Graham et al. 2016, p.686). McLachlan (2009) noted that a regression in learning after transition (leading to content repetition) both demotivates learners and fuels negative attitudes (p.202).

In interviews, such regression and repetition across transition points was raised by most teachers as a serious concern. Many reported that regardless of the level of Welsh attainment made by learners in primary school, their cluster secondary would take all learners back to the very start of the WLC:

> We used to have really good transition programmes with our secondary school, but we just find the secondary isn't interested in what we've been doing with the children at all. Regardless of how good they [the children] are at a certain subject, they're all just going to start at this point and either coast for year 7, or get challenged. Which is really unfortunate in terms of, we worked hard with these children. You know? And it's just we're not valued. We're not valued at all.
> *Interviewee 02*

> Our children leave us with, I think, excellent levels of Welsh. Unfortunately, then they go into the secondary system where they go back to counting to 10 and learning how to say their name, which is really frustrating.
> *Interviewee 04*

In many cases this 'back to square one' approach was attributed to a lack of assessment data that would inform more accurate targeting of content and pedagogy. As already mentioned, only one of the 81 participants reported their school passing on any data on Welsh language proficiency to their cluster secondary schools as part of transition. However, others who did report collecting data, suggested that secondary schools would ignore the Welsh language assessment information if it were provided:

> So, any assessment that we do, if we were to come up with a standard, it's not accepted anyway because they just kind of go, 'Well, we'll start again anyway because we have all these different people and different ones [assessments]'.
> *Interviewee 14*

> *Dw i'n meddwl mae nhw gwybod am y [asesiad] rhestr, ond i fod yn honest, pan mae plant yn gadael ni, mae nhw'n mynd i'r ysgol uwchradd a dechrau popeth eto. Reit nol i'r alphabet a 'pwy wyt ti?'*

> I think they know about the list [assessment], but to be honest, when children leave us, they go to the secondary school and start everything again. Right back to the alphabet and 'who are you?'
> *Interviewee 08*

So, we don't pass over assessment details, but they kind of do their own. We used to pass it [assessment data] over, but we realised they just do their own and they don't really want ours.
Interviewee 07

Some teachers ascribed this to a lack of trust in the levels allocated by primary teachers, which could be understandable given the variation and potential inaccuracy in existing protocols (see [section 4.1](#)). Others highlighted a more general lack of coordination and cooperation between primary schools and their cluster secondary schools:

IR: Would you say there's a good level of communication between the Welsh department of the secondary school and the cluster primaries?
IE: I would say that it's non-existent, if I'm honest… There is communication between myself and their ALNCo, but in regards to Welsh, no.
Interviewee 18

Yn anffodus, rydyn ni'n yn yr un safle gyda'r ysgol uwchradd, does dim lincs yna eto.

Unfortunately, we are in the same position with the high school, there are no links there yet.
Interviewee 08

Without clear communication between primary Welsh leads and secondary Welsh teachers, it is hard to see how learning continuity can be created for pupils in their Welsh language progression. Reflecting the research by McLachlan (2009), teachers attributed several negative impacts to this lack of continuity:

I've seen test scores from previous pupils that have gone shooting down [after transition] and I don't know how. So, I feel like, this student is set three when they shouldn't be. They weren't with me. So that continuity is broken. I think that that gap there, there's something wrong.
Interviewee 07

They're completely put off [the transitioning learners]. Because those that struggle anyway, they're already don't want to be taking part in it. And those that actually are quite good at it [Welsh] and achieving quite well, it's just boring for them. So, then the lessons become boring, because they've gone back to stuff they already know.
Interviewee 14

We've had some time to go into year 7 classes and observe how the children engage with Welsh and it's a bit of a mixture really, some children I feel disengage because it's too low level.
Interviewee 13

But what was happening was the enthusiasm for Welsh was just being destroyed because they were being told you need to go back to, 'Pwy wyt ti'. And then the enthusiasm for Welsh just sort of died, you know?
Interviewee 09

In contrast to these concerns about inter-institutional transitions, many respondents felt that intra-institutional transitions (between year groups or progression steps) were largely effective

and well-managed. Many reported well-established hand-over procedures that would include assessment, and discussions of learner progression/needs:

> The teacher that has a class in September can have a look [at the assessment data] where they were in the summer, decide what they need to plug whichever gaps they have to plug... there's always great baseline assessment against the key language patterns for the progression step, for the key stage 2 and all of the children have a baseline there... The teacher then pitches wherever it is she needs, or he needs, to start from there.
> *Interviewee 04*

However, some participants indicated that whilst these hand-over procedures existed for other subjects and areas (usually mathematics, English, and well-being), they did not include Welsh.

It appeared that proximity was an important factor in this far higher standard of internal transition. Primary teachers working on the same site and having established professional relationships appeared to make the transition process an integral part of school routine (Dunlop 2003). Conversely, the distance and lack of relationships with the secondary school was raised as an impediment to effective transition procedures.

4.4 Teacher attitudes to Welsh language assessment

In addition to questions concerning assessment provision and protocol, both the survey and interviews sought to gather data on teachers' attitudes towards the assessment of Welsh both generally and with regards to the specific approach of their individual institution. This is an important factor to consider, as attitudes have a significant impact on the way in which teachers' engage with assessment, impacting both the fidelity of assessment use, and the utilisation of assessment data (Summers 2023; Latif & Wasim 2022).

Two key themes were identified in the data collected: perceived importance of Welsh as a taught subject; and teachers' Welsh language competence/confidence. For clarity these elements will be explored individually below, although they obviously intersect to a certain degree.

4.4.1 Perceived importance of the Welsh language as a subject

The importance of Welsh and the status the language is afforded in the minds of learners, teachers and within institutional cultures, was a theme that was raised by almost every interview participant. Perceived importance is a well attested factor in learner motivation (Csizér & Dörnyei 2005), time allocation to instruction (Fitchett et al. 2014), and the status of different types of learning (Fullen 2016; Flores 2004). Given the variation in the degree to which learners, staff and institutions will internalise the Welsh language as part of their identity, a certain amount of inconsistency is unsurprising. In section 4.1 the relationship between the provision of assessment and the status of Welsh within the school was touched upon. Here the theme will be explored in more detail and from the opposite causative perspective (i.e. how teacher attitudes may impact on assessment, rather than the impact of assessment practice on attitudes).

This theme was not explicitly included in survey data and was largely raised independently by interviewees, though in the later stages of the study, as this theme became apparent, it was included in the interview schedule to ensure that a broad range of perspectives were available.

Given the position of the interviewees (current or previous participants on a WLS course), it is unsurprising that their personal attitudes to the language were overwhelmingly positive, seeing the Welsh language as an important element of English medium education. However, some reported more negative attitudes from colleagues and leadership:

> So, staff will flat out say to the children, 'I don't know Welsh'. And they wouldn't feel that teaching Welsh was their job necessarily. It wasn't part of their role... And I think that gives the attitude to the children then of. Well, Miss doesn't know it, Sir doesn't know it, so well, I'm not going to know it then, am I?
> *Interviewee 14*

> The culture is going to take a big change because there's not a lot of Welsh support by adults in the school [for Welsh]. But I think that's where myself and the Welsh lead have got to come together and start thinking about changes for the culture.
> *Interviewee 07*

This was not a ubiquitous feature, and other teachers reported that Welsh had a high status within their schools, often stemming from strong direction from the SLT:

> *Mae'r deputy head dod o Llanelli. Felly mae nhw'n hyderus i jyst 'have a go'. A dyma'r diwylliant o'r pennaeth i lawr rili.*

> The deputy head comes from Llanelli. So, they are confident to just have a go. And that's the culture from the head down really.
> *Interviewee 08*

> We do have a good ethos in this one. It is modelled that they hear the language on the corridor between the staff. There is Welsh music played regularly, so they understand that it is around them and it is a living thing.
> *Interviewee 16*

However, despite such examples, a positive school ethos with regards to Welsh cultural identity was not always mirrored by the Welsh language provision of the school:

> They've [the children] done the culture, so like I could show you their Welsh crafts that they did that were inspired by famous Welsh artists, the Eisteddfod [a festival of Welsh arts and culture] and that kind of thing. But Welsh language? They don't have like, workbooks or even resources around the classroom or anything like that. Yeah. It's just not a priority.
> *Interviewee 10*

This perspective of Welsh as having a lower priority than other subjects was reflected in the responses of participants. Teachers were often keen to clarify that staff had no personal opposition to the language, but that other areas of learning took precedence:

> I don't think it's deliberate. It's just, you have so much to do and I suppose if there's that one thing that you're going to forget is the thing that you're perhaps not so confident with. Yeah. I don't think it's. No, it's, it's not malicious or anything like that. I think it just is time pressure.
> *Interviewee 03*

It's not as if people are dead against it [Welsh] because they're not. They just don't have the skills... They don't hate Welsh or whatever, but they're just not confident. And because they don't get enough confidence, obviously they tend to shy away from it a bit more.
Interviewee 05

This link between teacher confidence in a subject and the priority allocated to it is consistent with studies in other subject areas such as maths (Lee 2005; Russo et al. 2020) and creative arts (Hennessy 2000). Respondents often reported that a common impact of such 'de-prioritisation' was Welsh being more likely to be dropped from the timetable that other subject areas, and that the inclusion of Welsh on the timetable was often seen as performative or perfunctory:

If the maths and the English skills have slipped, we need to sort of park Welsh a little bit and focus on them...I do feel that it can almost be the first thing to be pushed if the week is super busy. You never drop an English or a maths lesson, but you might go, 'Oh, I didn't quite get round to Welsh this week'.
Interviewee 06

[The attitude] was very much do Welsh when you want how you can. So, that's obviously the first thing in the timetable to go every time.
Interviewee 07

It's the first thing to go. It's the first thing. Oh, like, 'We haven't done Welsh, oh never mind'. It's very much that sort of attitude.
Interviewee 09

For the most part, interviewees did not attribute this to a lack of support amongst staff for the learning of Welsh, but rather as a response to extreme levels of time pressure on the weekly schedule. This pressure inevitably leading to some subjects being dropped in order to ensure 'priority' subjects (usually identified as mathematics and English) could be included. As already touched upon in section 4.1.1, some participants linked this prioritisation of maths and English with the assessment protocols of the school: whilst maths and English were assessed and monitored extensively, Welsh was often reported to be ignored or devalued from an assessment perspective:

This is why I do believe when you were talking about this as an idea to have an assessment system that is in place, it makes you accountable and it also gives you something to work towards. And then the children can see their own progress. I think something like this is essential. It doesn't exist in our school at the moment.
Interviewee 16

Welsh just isn't monitored like English and maths, and you know, it's completely natural that teachers prioritise the areas they know they'll be like held accountable for... if something is assessed, that it, it will find a spot on the timetable.
Interviewee 19

> If there was a test in writing skills and Welsh, and the children all failed the Welsh test but they did excellently on the English, I'd probably get a pat on the back... But if the English was poor and the Welsh was through the roof everyone would think, 'what's wrong with this teacher'.
> *Interviewee LA5*

This back wash effect from assessment, leading to the prioritisation of some subjects over others is supported by SLA and education literature generally (Fitchett et al. 2014; Cheng & Curtis 2004). Teachers feel that they will be penalised or held to account should learning outcomes in monitored subjects fall behind, but do not feel the same pressure in unassessed areas of learning. The impact of such accountability was also evident in the numerous referrals made by interviewees to Estyn (the education and training inspectorate for Wales) as a motivating factor for schools in terms of Welsh language teaching and provision:

> I think our head teacher is expecting an Estyn visit very soon. So, it's quite funny how that gives people focus. And they suddenly think, oh, good grief, it [Welsh] will be a recommendation. So, let's plough all our efforts, all our time into it. Let's get the ball moving.
> *Interviewee 01*

> They do obviously have an impact on Welsh and again I think the impact on Welsh ebbs and flows depending on when Estyn is coming in.
> *Interviewee 06*

> It's only really this year we started changing things because we had an Estyn inspection last year who have said, you know, Welsh isn't good enough. So, we've changed to it. Before, it was very much, do Welsh when you want however you can.
> *Interviewee 07*

> ...with Estyn putting it as a as a priority. Such a recommendation, it sort of opens more of a mindset of this. Like, this does need to become more propriety now, yes.
> *Interviewee 13*

This phenomenon of the reported de-prioritisation of Welsh appears therefore to be a compound effect of multiple factors: time-pressure, relatively subject status within school cultures, accountability and assessment practice.

School cultures are often self-perpetuating (Hinde 2004; Kaplan & Owings 2013). A relationship between accountability, subject status and teacher practice can therefore become cyclical: the lower subject status, resulting in de-prioritisation in teacher practice and CPD, leading to lower learner performance within the subject area, leading to further degradation in the subject status within the institution (Greany et al. 2018, p.14). It is possible that the lack of Welsh language assessment (and consequential perceived lack of accountability) reported by participants may be indicative of such a cycle in some schools. Paradoxically, such an effect is extremely difficult to establish statistically, as the lack of accountability is contingent on Welsh not being assessed, meaning no data is available. However, the prevalence of this theme in teacher interviews does raise the possibility that such a mechanism is at play.

Such cultural inertia within institutions can be difficult to change once established (Carrillo & Gromb 2007), but some teachers did report successful challenges to previous cultural norms surrounding Welsh within their schools:

> I do think it has become more prominent in the school. There is more of a drive that we should be embracing Welsh and using it as often as we can.
> *Interviewee 18*

> Welsh has not been at the forefront and you can tell when you go over there [to the school], but I think this year things have changed. This is why I wanted to do the course.
> *Interviewee 07*

Such examples demonstrate the impact school leadership and external motivators (e.g. inspections, CPD support, inter-institutional collaboration) can have on influencing the status of Welsh within individual institutions. However, whilst positive, such cultural shifts do not tend to address the apparent underlying cause of de-prioritisation: timetabling pressure.

4.4.2 Teacher competency in the Welsh language

A solution to the problematic time-pressures outlined above would be the integration of Welsh across the curriculum, ensuring learners have extensive exposure to the language throughout the school timetable, mitigating any impact of de-prioritisation. Such an approach is advocated in Content and Language Integrated Learning (CLIL) model (Coyle et al. 2010; Beard 2020) and the adapted Extensive Processing Instruction (EPI) approach (Smith & Conti 2023) which has been adopted by most consortia for CPD and forms the basis for the pedagogical input of the WLS courses. Teachers interviewed were largely positive about this approach, highlighting the importance it places on embedding patterns through revision, game-based pedagogy, and focus on oracy skills. However, teachers also raised the issue of staff competence as an impediment to the widespread adoption of this approach:

> It's another thing that's good on paper, isn't it? Basically, you say, 'Right, we're not going to do Welsh as a standalone lesson because we want children to speak Welsh all the time, not just in that lesson'... But then you've got a teacher that doesn't know Welsh and then you don't use it.
> *Interviewee 15*

> Yes, it's the language part that I think that the teachers aren't as confident in. Because there's, you know, there's a couple English teachers, there's a teacher who grew up in the US. So that's just, that's the struggle, specifically with the language.
> *Interviewee 10*

> *Dw i'n meddwl mae'n anodd i datblygu cynllun yn ysgol ble hanner o athrawon ddim yn gallu siarad Cymraeg o gwbl. Problem gyda'r cysondeb. Os dim ond un athro ddim yn gallu siarad Cymraeg, wedyn does dim cysondeb.*

> I think it's difficult to develop a plan in a school where half the teachers can't speak Welsh at all. A problem with consistency. If even one teacher can't speak Welsh, then there's no consistency.
> *Interviewee 11*

Whilst participants were confident that they themselves would be able to create such language rich environments, some highlight the disparity in their training, compared to the CPD available to their colleagues. Such authentic cross-curricular language integration is dependent to some extent on target language proficiency (Richards et al. 2013). Whilst the required level is not as demanding in the modified EPI approach as it would be for a full CLIL model, many respondents

highlighted the lack of Welsh language confidence/competence as a barrier to colleagues teaching the language:

> We've been kind of following the EAS packs, but like I said, it's too difficult and teachers don't really know. They don't have the skills to be able to take responsibility.
> *Interviewee 05*

> I think teachers are very scared of teaching Welsh because they don't really know what to do.
> *Interviewee 09*

These concerns regarding teacher confidence/competence often linked back to the topic of assessment. Within the CfW's delegated approach to assessment, schools and teachers are expected to create their own assessment tools, but many respondents felt they did not have the training or skills to develop Welsh language assessments. Interviewees highlighted the challenges faced by teachers in assessing an area in which they often had a low level of ability themselves:

> You need to be an expert in your field to come to convey confidence, you know. Potentially yes, I think it makes it significantly harder to confidently assess something, if you don't necessarily know what it is you're assessing.
> *Interviewee 12*

> I've had no training on developing assessments, let alone specific Welsh assessment. It feels like this has been plonked in our lap, on top of everything else, and we are just expected to know what to do.
> *Interviewee 19*

This is reflected in the survey data, which indicated only one of the 81 respondents had any kind of training on the assessment of Welsh specifically. Low levels of assessment literacy in teachers can lead to teachers either abstaining from assessment or producing assessments that are ineffective/misleading (Inbar-Lourie 2008; Weng & Shen 2022).

In addition to these concerns about more formal assessment tools, teacher proficiency was also discussed in the context of formative assessment practice:

> Because I don't have the knowledge myself, it would be very difficult for me to sense check the learners, and I do admittedly draw myself back from that sort of AFL then, because I don't feel qualified enough to make that judgement.
> *Interviewee 18*

> When you have a lack of confidence in something, especially language, immediately you're less comfortable in using familiar [AFL] strategies... I would feel genuinely nervous assessing a child, when I am not 100% sure that I'm doing it right. Because I don't want to assess someone and tell them, give them feedback, but the feedback is not entirely accurate. So, it definitely effects other staff.
> *Interviewee 20*

It is well established that target language proficiency has a significant impact on the accuracy, quality and quantity of formative feedback provided to learners (Fulcher 2012; Van Canh & Renandya 2017; Gebril & Brown 2019). This can lead to teachers providing untargeted or inappropriate feedback, potentially having a detrimental impact on learner progression. A lack

of target language proficiency can also result in teachers strictly adhering to prescriptive curricula content due to a lack of confidence in the language skills/knowledge (Gak 2011; Faour & Kiwan 2022), a theme that was picked up by a few of the interviewees:

> *Mae nhw'n jyst dilyn y scheme of work. Ti ddim yn gwybod sut i ddysgu plant y iaith, achos os ti ddim yn gwybod sut i siarad Cymraeg, wedyn ti ddim yn gwybod sut i ddysgu Cymraeg.*
>
> They just follow the scheme of work. You don't know how to teach children the language, because if you don't know how to speak Welsh, then you don't know how to teach Welsh.
> *Interviewee 15*

> IE: I don't know if other staff would be using as much formative assessment in their classes.
> IR: Why do you think of the staff members might use less?
> IE: I just think, I think because it's not as written down in schemes. If they're following it prescriptively it wouldn't necessarily to say, you know, 'try an activity on seesaw'.
> *Interviewee 13*

Such restricted provision-led practice can occur despite teachers understanding that a responsive approach to pedagogy and provision are beneficial to learners, and could contribute to the attitudinal factors explored in section 4.4.1.

4.5 Teacher recommendations and suggestions

Teachers were invited to make suggestions of what provision they felt would help them assess Welsh more effectively in their classrooms. The most common suggestions can be categorised into three areas: practicality, flexibility, and learning enhancement.

Practicality

Most teachers emphasised that any assessment provision for Welsh needed to be practical and easy to use. This included being time and resources efficient so as not to unduly impact on teaching and learning, not requiring a great deal of training or expertise to use, and not requiring complex record keeping or monitoring.

> You know, something that's quick and manageable, doesn't add to workload too much
> *Interviewee 03*

> I think you know with assessment sometimes it always becomes really time consuming and another extra job that a teacher just doesn't needs to have. They're very busy people.
> *Interviewee 06*

> In a dream world an app sort of situation, where the child can engage and press multiple choice or something, so that they could do it themselves.
> *Interviewee 16*

Aligned and Representative

Many teachers reflected that any provision would need to be adapted to the needs/level of their learners and align with their individual school curriculum:

It needs to have been covered, doesn't it? It needs to assess what's actually being taught. Isn't it? Because otherwise, what's the point?
Interviewee 05

My concern would be that any standard assessment would test them on things I've not taught them yet. You know, and every class is different. You'd need to be able to control the content, match it up with your lessons, or you may as well test their Japanese.
Interviewee 19

Some interviewees also contrasted an aspirational assessment with the current practices that only sampled a small number of selected children from each class, suggesting a more comprehensive or representative approach would improve the accuracy of data generated:

I would like something more reflective of the whole class of the school, not just cherry-picking students and listening to learners. There needs to be something broader.
Interviewee 07

A tool that the county can get an overview of could be very useful as someone who's now going to be monitoring Welsh a lot more within my school that would be useful.
Interviewee 14

As my focus would be, is the data accurate to that child? Have they actually showed where they are? Is that a true representation of where they're at right now?
Interviewee 16

Learning enhancement

Many teachers emphasised the need for any assessment provision to be focused on enhancing teaching and learning, rather than on producing arbitrary levels. Some were more specific, discussing the specific types of data which they would find useful in directing their teaching practice more effectively:

Something that will look at trends, at things that pupils are strong at, maybe sentence patterns that cohorts are weaker at, you know, looking at individuals, but then groups, you know, something that's usable.
Interviewee 01

You want to know the impact. That's really all I want. I only want to do things that are going to impact the learning, or my teaching, or my children. If it is just for the sake of doing an assessment exercise there is no point in my opinion.
Interviewee 06

It has to help learning. You know, it's all well and good, saying 'use this', but if you want teachers to use something, they have to find it useful. Or it just ends up in the back of a drawer.
Interviewee 19

Some teachers highlighted the lack of direct assessment feedback to learners as potential deficit in current approaches, advocating for tools that could facilitate learner directed development:

> I get told the feedback and then I implement it... There's no official feedback to them [the learners]. And when it comes to me, it may not necessarily be reflective of how it was at the time.
> *Interviewee 18*

> At the moment, any feedback is filtered through the teachers, it would be good for learners to have some ownership of their progression, to know how they're improving and what they need to do to develop further.
> *Interviewee 19*

5. Conclusion

This report aims to provide an overview into the Welsh language assessment practices prevalent in mainstream English medium primary schools in Wales. The key objectives of the study were to explore: the current formative and summative Welsh language assessment practices and tools in use; the extent to which teachers feel these practices and provisions meet the needs of teachers and pupils; and how the assessment of Welsh language proficiency is used as part of transition to ensure continuity of learning. The study also provided the opportunity for teachers to feedback on how they believed Welsh language assessment provision could be enhanced within their settings.

The data collected is reflective of a wide range of institutions, including schools from every region/consortium, and a wide range of socio-economic contexts.

5.1 Summary of the findings

The participants reported that most schools provide no assessment provision for Welsh. The provision that was provided in the minority of cases was often reported by teachers not to accurately reflect the proficiency of learners, provide limited actionable data for teachers, and increase workload. Where available, assessment tools also tended to have been developed in isolation, making them suitable for developing intra-institutional, but not inter-institutional understanding of learner competence and progression. This issue is compounded by the reported lack of cooperation within clusters concerning Welsh language standards and assessment. Very few schools currently provide any assessment information for Welsh as part of the transition between primary and secondary education, which several teachers in the study suggested leads to pupil demotivation and regression in learning. However, internal transitions are often reported to be handled well and to facilitate learning continuity effectively.

Formative assessment of Welsh is reported to be far more widespread that the use of formal assessment tools, though some participants indicate that a lack of teacher confidence in Welsh may have a negative impact upon the quality of feedback. Current formative assessment practices often appear to prioritise the assessment of short-term/immediate learning objectives rather than longer term learning progression. The perceived importance and status of Welsh within the school culture is reported to both impact, and be impacted by, the school's assessment practices and subject prioritisation.

5.2 Implications and recommendations

The findings of this study indicate that, in many schools, Welsh assessment that enhances long-term learning is not occurring. This raises a challenge for the EM primary education sector: how can the assessment of Welsh be developed and enhanced to improve learning outcomes

without placing additional pressure on teachers, and without undermining the principles and intentions of the CfW? Four key approaches may help address this challenge: developing teachers' Welsh language skills; improved provision for supporting teachers to assess Welsh effectively; greater communication and coordination between clusters and their associated secondary schools; and the enhancement of second language specific pedagogy.

The significant association between teacher language proficiency and the efficacy of formative assessment indicated in this study is consistent with the literature (Faez et al. 2021), demonstrating the on-going importance of upskilling the education workforce with regards to the Welsh language. Such improved teacher competence may also go some way to addressing the de-prioritisation of Welsh explored in section 2.4 (Sabokrouh & Barimani-Varandi 2013; Onodera 2025). Current initiatives such as the WLS courses, the subsidisation of Welsh course through the Canolfan Dysgu Cymraeg Genedlaethol [National Centre for Learning Welsh], and the CPD initiatives of regional consortia are likely to act as key facilitators of this skills development approach. A limiting factor highlighted by staff was the time pressure inherent in their work-life that acted as an impediment to developing their language skills. A greater allocation of curriculum time to Welsh language CPD would mitigate this problem, however there would obviously be significant cost implications for schools. It is encouraging that existing research (Tsang 2017; Faez et al. 2021) indicates that full fluency is not required to substantially improve the quality of teaching in a target language, making the aspiration to upskill teachers (to the extent that their practice is optimised) more attainable. A greater recognition of Welsh language competence as a professional attribute within the EM sector, enhancing the perceived status of the language and incentivising continued CPD within the workforce may also yield improved outcomes in this area (Badler et al. 2016).

The development of improved provision to support teachers in the assessment of Welsh offers an accessible and cost-effective way of improving the efficacy of assessment quickly (Sandefur 2022). Unlike the long-term development of teachers' language skills, the provision of tools to support the assessment of Welsh could have an immediate impact on teachers' ability to assess their learners' proficiency accurately, and communicate that proficiency effectively, thereby enhancing their pedagogy, content selection, remediation and pacing, whilst facilitating improved continuity of learning across transition points (Faxon-Mills et al. 2013). In addition, a tool that could be used across different institutions would provide a more objective understanding of learner performance, allowing teachers a more contextualised understanding of their learners' progression (Geranpayeh & Taylor 2013), and enabling empirical research (both academic and teacher-led 'action research') into the efficacy of different teaching methods and learning materials (Olsen et al. 2011).

There are two factors that potentially prevent the utilisation of a broad assessment provision: technical, and ideological. Technical challenges include alignment/comparability (i.e. how would a tool align with a school's Welsh curriculum and still provide comparability with a school using a different curriculum?), practicality (i.e. how would the assessment provide actionable data whilst not adding to teacher workload or timetabling), and accuracy (i.e. how would an assessment operationalise the key constructs of Welsh language competency given the restrictions of alignment and accuracy already mentioned). Perhaps a more fundamental challenge than these practical considerations is how any inter-institutional assessment could align with the principles of the CfW, which emphasizes the importance of contextualised assessment aligned with the school's socio-cultural make-up and environment, whilst retaining a sufficient level of standardisation to provide meaningful levels of comparability (WG 2024), as

highlighted in previous research (Sireci 2020; Skarpenes & Hidle 2024). A similar conflict exists between the need to easily communicate learner competence to facilitate learning continuity and the dangers of producing data that could be used to impose professional accountability based on learning outcomes (Cust-Hughes 2025). Despite these challenges, improved assessment provision offers the intriguing potential of providing teachers of all language proficiencies with feedback on learner development that could enhance their practice and the learning outcomes of pupils significantly (Vassiliou et al. 2023).

Greater communication and coordination within clusters and between cluster schools and the associated secondary schools offers the potential to address the damaging regression reported by the study participants (Cantley et al. 2021). This would provide primary schools with a better understanding of the expectations of learner competence by the transition point in year 6/7, allowing then to plan learning trajectories and modify content selection accordingly. It would also provide secondary schools with a better understanding of the proficiency of the learners moving into year 7, allowing them to adapt content and provision to ensure learning continuity is maintained, mitigating the risk of disengagement/demotivation associated with learning regression (Pfenninger & Lendl 2017). Such cooperation also offers the potential for the facilitation of staff CPD, with cluster secondary school well placed to contribute to both Welsh language and L2 pedagogy training to their cluster primary schools.

The specific skills and methods associated with language teaching were highlighted as a deficit in current ITE provision, with many teachers training outside of Wales and so receiving no guidance at all. Even those who trained in Wales who received instruction in Welsh as part of their training often report that the instruction was focused on learning the language, rather than language teaching pedagogy, a finding consistent with recent findings by Estyn (WG 2023). Providing training to teachers to enhance their specific L2 pedagogy may help improve the learning experiences of pupils, raise standards of Welsh in pupils, and develop the confidence of teachers leading to more favourable attitudes to the language. Many consortia already provide this training, and an expansion of these programmes alongside efforts from local authorities to encourage schools to take up these opportunities offers the potential to enhance teacher pedagogy and pupils' learning outcomes. More specific training around the use of formative assessment in L2 learning contexts, and the design/deployment of effective summative assessment for SLA would also likely have a significant impact on the provision and efficacy of Welsh language assessment (Karaman 2021; Anh 2023; Anamalai & Yatim 2019).

5.3 Recommendations for further research

It must be noted that the scope of the report does not include the perspectives of schools' leadership teams or consortia. This was a conscious decision, as the intention of this report was to gain an insight into the current assessment practice in schools as experienced by classroom teachers. Approaching school leaders or consortium representatives, whilst providing a legitimate perspective, was likely to result in data filtered through their experiences and priorities, rather than direct testimony from teachers which potentially provides more actionable insights (Briggs & Coleman 2019).

Now that such a practitioner focused perspective has been explored, the integration of school leadership and consortia perspectives would help contextualise the practice reported in this study, and inform potential approaches to address the challenges identified. Finally, a more detailed exploration of formative assessment practice, including direct lesson observations,

would provide a more accurate and nuanced picture of how AfL strategies are deployed in Welsh language lessons, and how they target and affect short and long-term learning goals.

Bibliography

Anamalai, T.R. and Yatim, M.H.M. 2019. A comparative study of formative assessment tools. *Journal of Information System and Technology Management, 4*(14), pp. 61-71.

Andrade, H. and Valtcheva, A. 2009. Promoting learning and achievement through self-assessment. *Theory into practice*, *48*(1), pp. 12-19.

Andrade, H. and Cizek, G.J. eds., 2010. *Handbook of formative assessment*. Oxford: Routledge.

Anh, T.T.N., Nguyen, N.T. and Nguyen, A.T.T. 2023. The impact of online self-assessment on learning outcomes and self-assessment skills among grade 11 students in Vietnam. *International Journal of Learning, Teaching and Educational Research, 22*(4), pp. 21-35.

Antibi, A. 2003. La constante macabre. *Ou comment a-t-on découragé des géné*.

Bachman, L.F. and Palmer, A.S. 1996. *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.

Ball, S.J. 2003. The teacher's soul and the terror of performativity. Journal of Education Policy, 18(2), pp. 215-228.

Beard, A. 2020. The suitability of Welsh language provision in English-medium schools to produce speakers of the language. *Wales Journal of Education*, *22*(2).

Bennett, R.E. 2011. Formative assessment: A critical review. *Assessment in education: principles, policy & practice*, *18*(1), pp. 5-25.

Berliner, D.C. and Nichols, S.L. 2007. *Collateral damage: How high stakes testing corrupts America's schools*. Cambridge, MA: Harvard Education Press.

Black, P. and Wiliam, D. 1998. Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), pp. 7–74.

Black, P. and Wiliam, D. 2009. Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of personnel evaluation in education)*, *21*, pp. 5-31.

Bleazby, J. 2015. Why some school subjects have a higher status than others: The epistemology of the traditional curriculum hierarchy. *Oxford Review of Education*, *41*(5), pp. 671-689.

Boud, D. and Falchikov, N. 1989. Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher education*, *18*(5), pp. 529-549.

Bradler, C., Dur, R., Neckermann, S. and Non, A. 2016. Employee recognition and performance: A field experiment. *Management Science, 62*(11), pp. 3085-3099.

Briggs, A. and Coleman, M. 2019. Research methodology in educational leadership and management. In: *Oxford Research Encyclopedia of Education*. Available at: https://oxfordre.com/education/view/10.1093/acrefore/9780190264093.001.0001/acrefore-9780190264093-e-606 [Accessed 15th April 2025]

British Academy. 2024. *British Academy response to the Department for Education's Curriculum and Assessment Review Call for Evidence: General response*. Available at: British-Academy-Cross-discipline-Curriculum-Review-response.pdf [Accessed 15th May 2025]

Cantley, I., O'Meara, N., Prendergast, M., Harbison, L. and O'Hara, C. 2021. Framework for analysing continuity in students' learning experiences during primary to secondary transition in mathematics. *Irish Educational Studies, 40*(1), pp. 37-49.

Carless, D. 2005. Prospects for the implementation of assessment for learning. *Assessment in Education: Principles, Policy & Practice*, 12(1), pp. 39–54.

Carney, E.A., Zhang, X., Charsha, A., Taylor, J.N. and Hoshaw, J.P. 2022. Formative assessment helps students learn over time: Why aren't we paying more attention to It?. *Intersection: A Journal at the Intersection of Assessment and Learning*, 4(1), p.n1.

Carrillo, J.D. and Gromb, D. 2007. Cultural inertia and uniformity in organizations. *The Journal of Law, Economics, & Organization*, 23(3), pp. 743-771.

Cheng, L. and Curtis, A. 2004. Washback or backwash: A review of the impact of testing on teaching and learning. *Washback in language testing*, pp. 25-40.

Coffey, J.E., Hammer, D., Levin, D.M. and Grant, T. 2011. The missing disciplinary substance of formative assessment. *Journal of research in science teaching*, 48(10), pp. 1109-1136.

Connor, C.M., Morrison, F.J., Fishman, B.J., Schatschneider, C. and Underwood, P. 2007. Algorithm-guided individualized reading instruction. *Science*, 315(5811), pp. 464-465.

Coyle, D., Hood, P. and Marsh, D. 2010. *CLIL: Content and language integrated learning* (Vol. 1). Cambridge: Cambridge university press.

Csizér, K. and Dörnyei, Z. 2005. The internal structure of language learning motivation and its relationship with language choice and learning effort. *The modern language journal*, 89(1), pp. 19-36.

Cust-Hughes, A. 2025. 'Datafication': The tension between accountability and the 'ethic of care' in primary teaching. *Impact* (2514-6955), (23). Available at: https://my.chartered.college/impact_article/datafication-the-tension-between-accountability-and-the-ethic-of-care-in-primary-teaching/ [Accessed 29th May 2025]

Day, C. and Gu, Q. 2010. The new lives of teachers. Oxford: Routledge.

Dunlop, A. 2003. Bridging Children's Early Education Transitions through Teacher Collaboration. NZARE. Available at: https://strathprints.strath.ac.uk/3339/1/strathprints003339.pdf [Accessed 14th April 2025]

Ellis, R., Loewen, S., & Erlam, R. 2006. Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, 28(2), 339–368.

Evangelou, M., Taggart, B., Sylva, K., Melhuish, E.C., Sammons, P. and Siraj-Blatchford, I. 2008. What makes a successful transition from primary to secondary school?. The Department for Children, Schools and Families, London, UK. Available at: https://webarchive.nationalarchives.gov.uk/ukgwa/20130401151715/https://www.education.gov.uk/publications/eOrderingDownload/DCSF-RR019.pdf [Accessed 12th February 2025]

Ewing, R. 2010. *The arts and Australian education: Realising potential*. Camberwell, VIC: Australian Council for Educational Research.

Karaman, P. 2021. The effect of formative assessment practices on student learning: A meta-analysis study. *International journal of assessment tools in éducation, 8*(4), pp. 801-817.

Kiwan, M. 2022. Change in Mathematics Teachers' Reliance on Teaching Resources in Response to COVID-19 Schooling Shifts. *Educational Research Journal- Lebanese University*, (28), 17.

Faez, F., Karas, M., and Uchihara, T. 2021. Connecting language proficiency to teaching ability: A meta-analysis. *Language Teaching Research, 25*(5), pp. 754-777.

Faxon-Mills, S., Hamilton, L.S., Rudnick, M. and Stecher, B.M. 2013. *New Assessments, Better Instruction?: Designing Assessment Systems to Promote Instructional Improvement*. Cambridge: RAND

Fitchett, P.G., Heafner, T.L. and VanFossen, P. 2014. An analysis of time prioritization for social studies in elementary school classrooms. *Journal of Curriculum and Instruction*, *8*(2), pp. 7-35.

Flores, M.A. 2004. The impact of school culture and leadership on new teachers' learning in the workplace. *International Journal of Leadership in Education*, *7*(4), pp. 297-318.

Foster, H. 2024. The impact of formative assessment on student learning outcomes: A meta-analytical review. *Academy of Educational Leadership Journal, 28*(S1), pp. 1-3.

Fulcher, G. 2012. Assessment literacy for the language classroom. *Language assessment quarterly*, *9*(2), pp. 113-132.

Fullan, M. 2016. *The new meaning of educational change*. London: Teachers college press.

Gaffney, C. 2018, November. Understanding the causes of inaccurate self-assessments: Extraversion's role. In *Proceedings of the 42nd annual Boston University Conference on Language Development*, pp. 238-251.

Gak, D.M. 2011. Textbook-An important element in the teaching process. *Hatchaba Journal*, *19*(2), pp. 78-82.

Galton, M., Gray, J. and Rudduck, J. 2003. Transfer and transitions in the middle years of schooling (7-14): Continuities and discontinuities in learning. *London: DfES*.

Gebril, A. and Brown, G.T. 2019. Quality of feedback in TESOL: A learning-oriented assessment perspective. *Quality in TESOL and Teacher Education*, pp. 140-149.

Geranpayeh, A. and Taylor, L. 2013. *Examinging Listening: Research and practice in assessing second language listening*. Cambridge: Cambridge University Press.

Glaser, R., Chudowsky, N. and Pellegrino, J.W. 2001. *Knowing what students know: The science and design of educational assessment*. Washington: National Academies Press.

Graham, S., Courtney, L., Tonkyn, A., & Marinis, T. 2016. Motivational trajectories for early language learning across the primary–secondary school transition. *British Educational Research Journal*, 42(4), pp. 682–702.

Graham, S., Courtney, L., Marinis, T. and Tonkyn, A. 2017. Early language learning: The impact of teaching and teacher factors. *Language Learning*, *67*(4), pp. 922-958.

Greany T, Cordingley P, Crisp B, Seleznyov S, Bradford M, Perry T. 2018. *Developing great subject teaching: Rapid evidence review of subject-specific continuous professional development in the UK*. London: Wellcome Trust.

Heine, S.J. and Hamamura, T. 2007. In search of East Asian self-enhancement. *Personality and Social Psychology Review*, *11*(1), pp. 4-27.

Hennessy, S. 2000. Overcoming the red-feeling: The development of confidence to teach music in primary school amongst student teachers. *British Journal of Music Education*, *17*(2), pp. 183-196.

Heritage, M. 2007. Formative assessment: What do teachers need to know and do?. *Phi Delta Kappan*, *89*(2), pp. 140-145.

Inbar-Lourie, O. 2008. Constructing a language assessment knowledge base: A focus on language assessment courses. *Language testing*, *25*(3), pp. 385-402.

Jindal-Snape, D. and Foggie, J. 2008. A holistic approach to primary—secondary transitions. *Improving Schools*, *11*(1), pp. 5-18.

Kaplan, L.S. and Owings, W.A. 2013. *Culture re-boot: Reinvigorating school culture to improve student outcomes*. London: Corwin Press.

Kowalski, F.V., Kowalski, S.E., Colling, T.J., Cuba, J.G., Gardner, T.Q., Greivel, G., Palou, E. and Russell, T. 2015. Using InkSurvey with pen-enabled mobile devices for real-time formative assessment II. Indications of effectiveness in diverse educational environments. *The impact of pen and touch technology on education*, pp. 307-314.

Kruger, J. and Dunning, D. 1999. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, *77*(6), p. 1121-1134.

Lee, J. 2005. Correlations between kindergarten teachers' attitudes toward mathematics and teaching practice. *Journal of Early Childhood Teacher Education*, 25(2), 173e184

León, S.P., Pantoja Vallejo, A. and Nelson, J.B. 2021. Variability in the Accuracy of Self-Assessments among Low, Moderate, and High Performing Students in University Education. *Practical Assessment, Research & Evaluation*, *26*, p.16.

Li, Y. and Zhang, G. 2016. Native or Non-Native-Speaking Teaching for L2 Pronunciation Teaching?--An Investigation on Their Teaching Effect and Students' Preferences. *English Language Teaching*, *9*(12), pp. 89-97.

Liu, R. 2018. Gender-math stereotype, biased self-assessment, and aspiration in STEM careers: The gender gap among early adolescents in China. *Comparative Education Review*, *62*(4), pp. 522-541.

MacIntyre, P.D., Noels, K.A. and Clément, R. 1997. Biases in self-ratings of second language proficiency: The role of language anxiety. *Language learning*, *47*(2), pp. 265-287.

Marsh, H.W. 1987. The big-fish-little-pond effect on academic self-concept. *Journal of educational psychology*, *79*(3), pp. 280-295.

Miller, E. 2025. Internal assessment and moderation in schools: Getting it right. Available at: [Internal teacher assessment and moderation in schools](#) [Accessed 27 March 2025]

Olsen, R.B., Unlu, F., Price, C. and Jaciw, A.P. 2011. *Estimating the Impacts of Educational Interventions Using State Tests or Study-Administered Tests*. National Center for Education Evaluation and Regional Assistance. Avialable at: https://files.eric.ed.gov/fulltext/ED524806.pdf [Accessed 15th May 2025]

Onodera, T. 2025. Are foreign language teaching enjoyment and motivation two sides of the same coin?. *International Review of Applied Linguistics in Language Teaching, 63*(1), pp. 759-782.

Pfenninger, S.E. and Lendl, J. 2017. Transitional woes: On the impact of L2 input continuity from primary to secondary school. *Studies in Second Language Learning and Teaching*, *7*(3), pp. 443-469.

Quick, L. 2024. *The SATs Effect: The verdict from year 6 teachers*. Available at: https://www.morethanascore.org.uk/wp-content/uploads/2024/05/TheSATsEffect.pdf [Accessed 15th May 2025]

Richards, H., Conway, C., Roskvist, A. and Harvey, S. 2013. Foreign language teachers' language proficiency and their language teaching practice. *The Language Learning Journal*, *41*(2), pp.231-246.

Russell, P.O. 2024. *Automagic Pilot - Summary Report: A process evaluation of the Automagic Welsh language intervention pilot conducted September 2023 to July 2024.* Available at: https://orca.cardiff.ac.uk/id/eprint/174145/1/Automagic%20Pilot%20-%20Summary%20Report%20.pdf [Accessed 27th May 2025]

Russo, J., Bobis, J., Sullivan, P., Downton, A., Livy, S., McCormick, M. and Hughes, S. 2020. Exploring the relationship between teacher enjoyment of mathematics, their attitudes towards student struggle and instructional time amongst early years primary teachers. *Teaching and Teacher Education*, *88*, 102983.

Sabokrouh, F. and Barimani-Varandi, S. 2013. The effect of EFL teachers' attitude toward English language and English language proficiency on their sense of efficacy. *Journal of Advances in English Language Teaching, 1*(4), pp. 117-125.

Sandefur, J., 2022. *Schooling for All: Feasible Strategies to Achieve Universal Education.* Center for Global Development. Available at: https://www.cgdev.org/sites/default/files/schooling-for-all-feasible-strategies-universal-eduction.pdf [Accessed 28th May 20225]

Schildkamp, K., van der Kleij, F.M., Heitink, M.C., Kippers, W.B. and Veldkamp, B.P. 2020. Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International journal of educational research*, *103*, p.101602.

Schmidt, F.T., Kaiser, A. and Retelsdorf, J. 2023. Halo effects in grading: An experimental approach. *Educational Psychology*, *43*(2-3), pp. 246-262.

Sireci, S.G. 2020. Standardization and UNDERSTANDardization in educational assessment. *Educational Measurement: Issues and Practice*, 39(3), pp. 100-105.

Skarpenes, O. and Hidle, K.M.W. 2024. The clash of cultures: Individualization and standardization in education. *Policy Futures in Education, 22*(7), pp. 1387-1403.

Smith, S and Conti, G. 2023. *The Language Teacher Toolkit*. Ho Chi Minh City: Piefke Trading.

Summers, D. 2023. Teachers' use of assessment data to improve student achievement. *Literature Reviews in Education and Human Services*, *2*(2), pp. 21-49.

Torrance, H. 2012. Formative assessment at the crossroads: Conformative, deformative and transformative assessment. *Oxford Review of Education*, *38*(3), pp. 323-342.

Tóth, E. and Csapó, B. 2022. Teachers' beliefs about assessment and accountability. *Educational Assessment, Evaluation and Accountability*, *34*(4), pp. 459-481.

Tsang, A. 2017. EFL/ESL teachers' general language proficiency and learners' engagement. *RELC journal, 48*(1), pp. 99-113.

Van Canh, L. and Renandya, W.A. 2017. Teachers' English proficiency and classroom language use: A conversation analysis study. *RELC journal*, *48*(1), pp. 67-81.

Van Nes, F., Abma, T., Jonsson, H. and Deeg, D. 2010. Language differences in qualitative research: is meaning lost in translation?. *European journal of ageing*, *7*, pp. 313-316.

Vassiliou, S., Papadima-Sophocleous, S. and Giannikas, C.N. 2023. Technologies in second language formative assessment: A systematic review. *Research Papers in Language Teaching and Learning, 13*(1), pp.50-63.

Welsh Government. 2019. *Statutory assessment arrangements for the Foundation Phase and end of Key Stages 2 and 3*. Available at: https://hwb.gov.wales/api/storage/53889241-676d-4e36-8e69-b9129f8ccf51/statutory-assessment-arrangements-for-the-foundation-phase-and-end-of-key-stages-2-and-3-190909.pdf [Accessed 5th February 2025]

Welsh Government. 2021. *Guidance on school categories according to Welsh-medium provision*. Available at: https://www.gov.wales/sites/default/files/publications/2021-12/guidance-on-school-categories-according-to-welsh-medium-provision.pdf [Accessed 15th May 2025]

Welsh Government. 2022. The Transition from Primary to Secondary School (Wales) Regulations 2022. Available at: The Transition from Primary to Secondary School (Wales) Regulations 2022 [Accessed 10th March 2025]

Welsh Government. 2023. *Support for Welsh in Initial Teacher Education: What we plan to do in response to Estyn's report on providing support to develop student teachers' Welsh language skills*. Available at: https://www.gov.wales/sites/default/files/pdf-versions/2023/10/5/1696592469/support-welsh-initial-teacher-education.pdf [Accessed 1st June 2025]

Welsh Government. 2024. Supporting learner progression assessment guidance. Available at: https://hwb.gov.wales/curriculum-for-wales/assessment-arrangements/supporting-learnerprogression-assessment-guidance [Accessed 13th December 2024]

Weng, F. and Shen, B. 2022. Language assessment literacy of teachers. *Frontiers in Psychology*, *13*, 864582.

Wiliam, D. 2011. *Embedded Formative Assessment*. Bloomington: Solution Tree Press.

Wiliam, D. 2018. *Creating the schools our children need: Why what we're doing now won't help much (and what we can do instead)*. West Palm Beach, FL: Learning Sciences International.

Yusuf, F.N. and Novita, O.E. 2020. EFL teachers' perceived language proficiency and teaching effectiveness. *Indonesian Journal of Applied Linguistics*, 9(3), pp. 580-588.

Appendices

Item 1 – A cross tabulation of teacher language proficiency (expressed as a binary: A2 and lower / B1 and higher) and teacher self-appraisal of their familiarity with their students' proficiency in Welsh (expressed as a binary: less than moderately well / moderately well or better) subjected to a chi2 test for significance.

```
    RECODE of
 Welsh_Ability
    (How would        RECODE of
you rate your   Know_own_pupils (How
        Welsh   well do you feel you
     language    know your learners'
     speaking         individu
      ability │ Not well  Moderatel │     Total
──────────────┼──────────────────────┼──────────
 A2 and Lower │      17         17   │        34
              │   50.00      50.00   │    100.00
──────────────┼──────────────────────┼──────────
B1 and Higher │      11         34   │        45
              │   24.44      75.56   │    100.00
──────────────┼──────────────────────┼──────────
        Total │      28         51   │        79
              │   35.44      64.56   │    100.00

      Pearson chi2(1) =   5.5279   Pr = 0.019
```

Item 2 – A logistic regression showing the odds ratio of teacher self-assessing as knowing their learners' proficiency moderately well or better for different levels of self-assessed teacher language proficiency, and whether they carry out formative assessment in addition to any formal provision provided by the school.

```
Logistic regression                              Number of obs =      78
                                                 LR chi2(2)    =   13.84
                                                 Prob > chi2   =  0.0010
Log likelihood = -44.002207                      Pseudo R2     =  0.1359
```

Know_own_pupils_Bi	Odds ratio	Std. err.	z	P>|z|	[95% conf. interval]	
Welsh_Ability	2.118408	.5570794	2.85	0.004	1.265227	3.546915
Additional_form_ass	3.346644	1.784089	2.27	0.023	1.17716	9.514443
_cons	.0232606	.0309622	-2.83	0.005	.0017123	.3159764

Note: _cons estimates baseline odds.