

Japanese and Welsh Evaluation Culture in Dialogue

Reaching into Research Article for the Association for the Study of Primary Education (ASPE)

Williamson-Dicken, M., Blackmore, R., Childs, D., Phillips, N., Powles, A., Soper, C. & Stokes, M. (2025)

Abstract

This paper compares Tokyo's school-led evaluation framework, as guided by Japan's Ministry of Education, Culture, Sports, Science and Technology (MEXT), with Wales's reformed Estyn inspection model of 2024. Drawing on desk-based research and the fieldnotes of an eight-day visit by Welsh immersion specialists to Tokyo schools, it explores how Tokyo evaluates immersion teaching quality, assesses learner outcomes, organises feedback loops and professional learning, and negotiates the tension between national standardisation and programme-specific flexibility. By contrasting these practices with Estyn's external inspection cycle, which is organised around teaching and learning, wellbeing and support, and leadership and improvement, the analysis identifies convergences and divergences that suggest practical adaptations for Welsh immersion education. In particular, the paper proposes that Welsh-medium settings might deepen their culture of self-evaluation, formalise mechanisms for learner feedback, pilot Japanese-style lesson study, integrate CEFR benchmarks into local assessment and invite low-stakes external review to reinforce Estyn's accountability processes.

Introduction

Education systems worldwide strive to balance accountability, professional growth and responsiveness to local contexts. Language immersion programmes present particular challenges because schools must ensure that learners are developing high-level proficiency in a second language so that they can equitably access the curriculum. In October 2025, a delegation from Torfaen County Borough Council, comprising of senior officers, a headteacher, immersion specialists and governing-body representatives, spent eight days in Tokyo observing immersion provision and analysing the city's evaluation culture. Their report revealed a model built on annual school-led self-review, transparent publication of findings and validation by community panels, set against Wales's system of external inspection by Estyn.

Since the early 2000s, Japan has shifted from informal internal reviews to a structured system of self-evaluation mandated by national policy. The 2006 School Evaluation Guidelines and the 2007 amendment to the School Education Law require each school to establish clear targets, conduct annual self-evaluations, publish reports and implement improvement measures. Local external panels drawn from parents, community members and academic experts validate each school's self-assessment and thereby foster greater transparency and shared responsibility for school improvement.

In Tokyo's immersion programmes, whether designed to boost English proficiency among Japanese learners or to integrate non-native speakers into Japanese-medium instruction, evaluation relies on multi-dimensional criteria. These criteria cover the integration of content and language objectives, dual measures of academic and language achievement, learner wellbeing, differentiation strategies and the creative use of community and cultural resources. Professional learning is supported by the *jugyō kenkyū* or lesson-study model, in which teachers collaboratively plan lessons with clear objectives, observe each other's practice and refine their approaches through iterative cycles of shared reflection.

By comparison, Wales's inspection framework as reformed in 2024 is organised around three inspection areas: teaching and learning; wellbeing, care, support and guidance; and leadership and improvement. Estyn's inspectors, including Her Majesty's Inspectors, peer inspectors and lay inspectors, observe lessons, interview staff, learners and parents, and review documentation before publishing narrative reports that highlight each school's strengths and areas for development. Although schools in Wales prepare self-evaluation reports, it is the external inspection that carries the primary evaluative weight.

This paper explores four central questions. Firstly, what criteria underpin Tokyo's evaluation of immersion teaching quality? Secondly, which methods and metrics inform the assessment of immersion learner outcomes? Thirdly, how does Tokyo organise its feedback loops and professional learning to foster teacher growth and classroom innovation? Finally, how does Tokyo reconcile national standardisation with the distinctive needs of immersion programmes? By mapping Tokyo's practices against Estyn's framework, the paper identifies adaptations that can strengthen Welsh immersion education and cultivate a culture of continuous school-led self-improvement.

Literature Review

The shift in Japan towards a structured school-led evaluation culture began in the early 2000s as part of broader reforms aimed at enhancing decentralised accountability and quality assurance. The 2006 School Evaluation Guidelines issued by MEXT provided exemplar indicators for evaluating curriculum, pedagogy, learner support, safety and community engagement. One year later the School Education Law was amended to mandate annual self-evaluations

and improvement measures for all schools. Scholars have described this transition as introducing an evaluative-state model under which schools themselves assume primary responsibility for assessing and improving their performance rather than relying solely on external inspections.

Immersion education adds complexity to this evaluative challenge because academic attainment, language proficiency and cultural integration must all be measured together. Lee and Niiya (2021) have emphasised the importance of robust language proficiency assessment that aligns with Common European Framework of Reference for Language (CEFR) to support migrant integration and track immersion progress. Wales's Welsh in Education Bill (2025) similarly responds to this need by mandating CEFR benchmarks for both Welsh and English language assessment in bilingual settings. This is expected to be in place by 2030.

Professional learning underpins effective evaluation and improvement. The *jugyō kenkyū* or lesson-study model involves teacher teams collaboratively planning a research lesson, observing its delivery and engaging in reflective analysis to refine practice. This approach nurtures deep professional dialogue, collective problem solving and incremental innovation. In Wales, professional learning communities exist and systems such as triad professional learning are adopted by some settings, but systematic lesson-study practices are far less established, indicating a clear opportunity for cross-cultural adaptation.

Empirical case studies illustrate the efficacy of integrated evaluation and collaborative professional learning in immersion contexts. Shima (2015) found in a study of an English immersion programme in Shizuoka that combining Content and Language Integrated Learning with ongoing peer review led to significant gains in both subject mastery and language proficiency. The Asia Society (2021) reported that Japanese local authorities engaging community stakeholders through evaluation panels secured stronger support and resources for immigrant and immersion programmes. At the same time Estyn's thematic review of Welsh education (2022) identified variability in self-evaluation quality and insufficient differentiation in evaluation criteria, underscoring the need for more rigorous programme-specific evaluation processes.

Taken together this literature highlights the importance of robust multidimensional self-evaluation cultures underpinned by collaborative professional inquiry and validated by community stakeholders in driving accountability and innovation in bilingual and immersion settings. The following sections describe Tokyo's model in detail, contrast it with Wales's inspection framework and propose targeted adaptations for Welsh immersion education.

Tokyo's Evaluation Framework

Tokyo's evaluation system is centred on the annual self-evaluation cycle mandated by MEXT. At the start of each academic year schools set clear context-specific targets drawn from MEXT's exemplar indicators. These targets cover curriculum coverage, learner progress, safety, wellbeing and community engagement. Throughout the year schools gather both quantitative and qualitative data against these targets and compile a self-evaluation report. The report is then published on the school's website to ensure transparency and accountability.

Local external panels comprising parents, community members and sometimes academic experts review each school's self-evaluation report. Panel members conduct site visits, interview staff, learners and parents and either validate or challenge the school's findings. Their advisory feedback is formally recorded and integrated into the school's improvement plan thereby fostering collective ownership of school development and ensuring that self-evaluation remains attuned to community expectations.

Within this school-led community-validated framework immersion teaching quality is assessed against criteria that extend MEXT's general teaching standards. Firstly, evaluators scrutinise curriculum integration to confirm that lesson objectives align simultaneously with subject content requirements and language development goals. Lesson plans are reviewed to ensure that teachers specify both content outcomes and language proficiency targets and that the required lesson hours are delivered in the immersion language. Lesson sequencing is examined to verify a coherent progression of subject matter and linguistic complexity.

Secondly, learner outcomes are measured through a dual assessment of academic performance and language proficiency. Schools employ commercial tests such as EIKEN for English immersion classes alongside bespoke in-house rubrics referencing CEFR descriptors for both English and Japanese. Academic achievement in mainstream subjects delivered through the immersion language is analysed alongside language assessment results to provide a comprehensive view of each learner's progress.

Thirdly, learner voice is embedded as an evaluation criterion. Structured surveys and focus groups capture pupil perceptions of comprehension, engagement and support in immersion lessons. These instruments prompt learners to rate the clarity of instruction, the appropriateness of pacing and the availability of scaffolding when complex language or concepts arise. Schools analyse this feedback to identify areas for improvement such as additional visual aids or alternative explanatory strategies and they incorporate the findings into their action plans.

Fourth, differentiation and scaffolding strategies are evaluated to ensure that teaching meets diverse learner needs. Observers look for evidence of personalised instruction through the use of visual aids, simplified language for beginners, peer grouping for mixed proficiency levels and extension activities for advanced learners. The aim is to assess how well teachers adjust content and language demands to each learner's profile.

Fifth, the use of community and cultural resources is taken into account. Schools document the involvement of native-speaker assistants, visits to cultural institutions, collaboration with local volunteers and participation in community events. Evaluators consider how such resources enrich immersion experiences by providing authentic language exposure and deepening learners' cultural understanding.

Immersion learners' outcomes are gauged through a combination of quantitative measures and qualitative indicators. Quantitative data include standardised language tests, termly internal proficiency checks tied to CEFR levels, academic scores in core subjects and transition rates from support classes to

mainstream settings. Qualitative measures encompass classroom observations of learner participation, narrative records of engagement in cultural and extracurricular activities, and interviews probing learners' sense of belonging and confidence in using the immersion language.

A distinguishing feature of Tokyo's outcome evaluation is its emphasis on growth trajectories rather than single-point scores. Evaluators track year-on-year improvements in both language proficiency and subject-matter achievement. They recognise that learners may face temporary dips in performance as they adapt to content delivered in a second language and they value continuous progress over isolated test results. This longitudinal perspective allows schools to contextualise early challenges and to validate the added value of immersion once learners stabilise and excel across both academic and linguistic domains.

Feedback and professional learning are organised through several interlocking mechanisms. Local external panels provide advisory reports that school leaders integrate into improvement priorities. The teacher appraisal system, administered by principals or senior colleagues, includes classroom observations focusing on language use, differentiation and learner engagement, followed by constructive feedback sessions designed to support teacher development.

At the heart of professional learning lies the *jugyō kenkyū* or lesson-study model. Immersion teachers participate in regular cycles of collaborative planning, observation and reflective analysis. A research lesson is co-designed with explicit content and language objectives and taught by one teacher while colleagues observe. After the lesson participants meet to analyse learner comprehension, language use, differentiation strategies and cultural integration activities. Through this collegial process teachers identify effective practices and areas for refinement. Lesson study embeds continuous improvement into daily routines and enables teachers to trial and refine immersion methodologies in a supportive environment.

At the system level MEXT promotes Plan-Do-Check-Act cycles and designates pilot "*kenshū*" schools to trial innovative approaches. Findings from these pilot schools are disseminated through professional development workshops, academic publications and inter-school visits. This ensures that successful practices scale across Tokyo's network of immersion programmes and that lessons learned in model schools inform broader policy and practice.

A central challenge in immersion evaluation is balancing national standardisation with programme-specific flexibility. Tokyo addresses this by treating MEXT's guidelines as non-binding exemplars and by inviting schools to formulate original targets and indicators reflecting their unique contexts. The tri-level evaluation model, comprising of self-evaluation, community panel validation and optional third-party expert review, maintains consistency in evaluation processes while permitting immersion schools to add or modify criteria to reflect bilingual objectives. When immersion learners show initial dips in mainstream subject scores evaluators interpret these results in context by examining whether learners catch up over time and whether school interventions have mitigated any long-term impacts. Conversely when immersion learners meet or exceed core academic benchmarks while achieving significant language gains evaluators recognise and commend the programme's holistic success. This flexible approach avoids penalising innovation while ensuring that fundamental academic standards remain inviolable.

Comparison with Estyn's Framework

Wales's Estyn inspection framework as reformed in 2024 shares Tokyo's ultimate aim of securing high-quality teaching and learning but operates through a markedly different structure. Estyn inspections occur on a multi-year cycle and are conducted by teams of professional and lay inspectors. These inspectors observe lessons, interview staff, learners and parents, and review documentation under three inspection areas: teaching and learning; wellbeing, care, support and guidance; and leadership and improvement. Reports are published publicly and provide narrative evaluations of significant strengths and areas for development. Estyn deliberately avoids assigning summative grades in order to encourage a focus on descriptive feedback and targeted recommendations.

Unlike Tokyo's school-led continuous self-evaluation Estyn's model places evaluative authority firmly in the hands of external inspectors. Although schools in Wales produce self-evaluation reports and use these to create development plans, it is the inspectorate's findings that drive ultimate accountability and improvement mandates. Estyn recommendations are not guidance only, they are mandatory improvement foci. Where as in Tokyo, any recommendations are guidance for school leadership.

Estyn places greater granularity on learner wellbeing and support under its second inspection area. Inspectors probe pupil voice, safeguarding measures, mental-health provision and inclusion practices in depth. While Tokyo's self-evaluations include aspects of learner guidance and safety they do not always measure the full spectrum of wellbeing indicators with the same level of detail. Conversely Tokyo's emphasis on harnessing community and cultural resources in immersion programmes exceeds the extent to which Estyn's framework foregrounds local cultural partnerships, despite Estyn's requirement to evaluate how schools promote Welsh cultural ethos and heritage across the curriculum. However, there is a danger with Tokyo's processes which sometimes see local biases and local politics given too much focus in the development plans.

Language assessment also differentiates the two systems. Estyn's inspectors systematically evaluate Welsh language development alongside English literacy across all schools. This practice reflects national priorities under *Cymraeg 2050* and ensures that immersion pedagogy is assessed by bilingual inspectors. In Tokyo, although immersion evaluations are increasingly aligned to CEFR descriptors, there remains significant variation across wards. This contrast suggests opportunities for reciprocal learning. Welsh immersion settings could adopt aspects of Tokyo's collaborative lesson study for professional development while Tokyo schools might benefit from Wales's systematic use of CEFR benchmarks in bilingual assessment.

Implications for Welsh Immersion Education

Tokyo's model underscores the value of embedding rigorous multidimensional self-evaluation cycles at school level validated by a variety of stakeholders and reinforced by collaborative professional learning. For Welsh immersion settings deepening self-evaluation culture could involve establishing annual cycles dedicated to immersion goals systematically gathering learner feedback on bilingual pedagogy through surveys and focus groups and engaging community

representatives in validation panels to co-construct development priorities. Such practices would complement Estyn's external inspection regime and empower schools to address immersion-specific challenges promptly and proactively.

Institutionalising Japanese-style lesson study within Welsh immersion contexts would create a sustained mechanism for professional inquiry. Teacher teams could co-design research lessons focused on vocabulary development scaffolding strategies or cultural competencies observe each other's practice and engage in reflective analysis to refine instructional approaches. This ongoing peer-driven model would enhance professional learning beyond the episodic nature of conventional inset training and foster a culture of shared responsibility for immersion success.

The systematic use of CEFR descriptors for both Welsh and English now mandated in national policy could be leveraged at school level to track learner progress more precisely inform differentiation strategies and align internal assessment practices with Estyn's evaluation criteria. Finally inviting low-stakes external review by immersion experts or immersion-experienced headteachers between formal inspections would offer developmental feedback disseminate best practice and sustain momentum for continuous improvement.

Recommendations

1. Increase Emphasis for a Systematic Self-Evaluation Culture
2. Formalise Learner Feedback Expectation Loops to Inform Self-Evaluation and Drive Change
3. Pilot Japanese Style Jugyō Kenkyū (Lesson Study) Improvement Mechanisms at a Micro and Meso Level
4. Pilot the use of the Common European Framework of Reference for Language (CEFR) in Assessment Processes at School Level (Micro and Meso) to Correspond to the Requirements of the Welsh Language in Education Bill (2025)
5. School Leaders to Incorporate More Informal External Perspective without High-Stakes Inspection to Support Self-Evaluation
6. Establish a More Nationalised Approach to What Constitutes Holistic Success and Measure this Consistently at Micro, Meso and Macro Level, Especially in Language Immersion

Conclusion

Tokyo's self-evaluation paradigm demonstrates how annual school-led review cycles validated by community panels and reinforced by collaborative lesson study can drive continuous improvement in immersion education. Its multidimensional criteria capture curriculum integration, language proficiency, cultural integration and learner wellbeing thereby offering a holistic blueprint for bilingual provision. Wales's Estyn inspection framework with its public narrative reports across teaching, wellbeing and leadership areas ensures transparent accountability and alignment with national priorities such as Cymraeg 2050. The 2024 shift towards descriptive evaluations without summative grades further aligns Estyn's ethos with Japan's commitment to improvement rather than sanction.

Welsh immersion education stands to gain by adopting key aspects of Tokyo's approach. Deepening self-evaluation culture, formalising learner-voice mechanisms, piloting lesson study leveraging CEFR benchmarks at school or centre level and instituting low-stakes external review will strengthen professional learning and foster resilient learner-centred bilingual communities. By combining robust internal reflection with structured external insight Welsh immersion settings can ensure that learners achieve academic excellence high-level bilingual proficiency and meaningful engagement with their cultural heritage.

References

- Ahn, R., Shimojima, Y., Mori, H. and Asanuma, S. (2018) 'Japan's innovative approach to professional learning', *Phi Delta Kappan*, 99(6), pp. 45–51.
- Asia Society (2021) *Multilingual Education in Japan: Policy Responses to Increasing Diversity*. New York: Asia Society Centre for Global Education.
- Central Council for Education (2017) *Report on the Vision for Education in the 21st Century: Toward the Realisation of a Curriculum Open to Society*. Tokyo: Ministry of Education, Culture, Sports, Science and Technology.
- Estyn (2024) *Inspection Framework for Education and Training Settings in Wales*. Cardiff: Estyn.
- Hashimoto, A. (2018) 'Note on the characteristics of Japanese school evaluation: Seven factors obstructing successful implementation', *Japanese Journal of Evaluation Studies*, 18(2), pp. 35–54.
- Katsuno, M. and Takei, T. (2008) 'School evaluation at Japanese schools: Policy intentions and practical appropriation', *London Review of Education*, 6(2), pp. 171–181.
- Lee, J. and Niiya, Y. (2021) *Language Proficiency and Migrant Integration in Japan*. Tokyo: Ministry of Education, Culture, Sports, Science and Technology.
- Ministry of Education, Culture, Sports, Science and Technology (MEXT) (2006) *School Evaluation Guidelines*. Tokyo: MEXT.
- MEXT (2007) *School Education Law* (Article 42). Tokyo: MEXT.
- MEXT (2024) *Results of School Evaluation and Meta-Evaluation Studies: Trends and Challenges in Quality Assurance*. Tokyo: MEXT.

Project Tokyo (2025) *Project Tokyo: Japanese and Welsh Evaluation Culture in Dialogue*. Torfaen: Carreg Lam.

Shima, K. (2015) *Curriculum Initiative Report: The Impact of an Immersion Programme on English Language Teaching in Japan*. Kanazawa: Foreign Language Institute, Kanazawa University.

Yomiuri Shimbun (2024) 'Writing skills among junior high students declining; urgent action needed', *Yomiuri Shimbun* (online). Available at: <https://www.yomiuri.co.jp> (Accessed: 28 October 2025).