

Beyond a checklist? A multiple case study exploration of the perceptions and experiences of teachers, mathematics subject leaders and senior leaders of the evaluation of primary mathematics teaching.

Dr Sam Parkes

University of Reading

Whilst evaluation of primary mathematics teaching is a well-embedded element of school improvement practice, the perceptions and experiences of those involved are poorly represented and under-researched both empirically and theoretically. This study offers new empirical data on current evaluation processes in primary mathematics. It also offers a new theorisation of the evaluation of primary mathematics teaching through its use of a conceptual framework focusing on professional development, professional knowledge, and professional identity. Key findings highlighted inconsistencies of perception and experience in relation to effective mathematics teaching, knowledge of primary mathematics and evaluation processes, and clarity of purpose and ownership of evaluations. These variations support the conclusion that there is a need for fairer, more coherent, and more useful evaluation processes of primary mathematics teaching and a new model for the provision of these based on mutual engagement, joint enterprise and a shared repertoire of tools is proposed.

Keywords: evaluation; primary; professional; teaching

The Context and the Problem(s)

The overarching aim of this study was to better understand the evaluation of primary mathematics teaching through the perceptions and experiences of both evaluators and the evaluated. Four ‘problems’ were identified as part of the contextual landscape in the UK: Problem 1: A potential lack of shared understanding of the varied purposes and consequences of evaluations; Problem 2: The impact of significant changes in mathematics curriculum, assessment and pedagogy on the evaluation process; Problem 3: Under-representation of key stakeholder voices in research and lack of clarity around the relationship between their roles; Problem 4: The risk of de-professionalisation of teachers and leaders through the evaluation process

Conceptual Framework and Research Questions

Following the identification of these problems, a conceptual framework guided the study to systematically elucidate understanding of evaluating primary mathematics teaching. This framework consisted of three inter-related concepts chosen to provide new theoretical insights: professional development (Wenger, 1999; Knowles et.al., 2020; Levin-Rozalis et.al., 2015), professional knowledge (Rowland et.al., 2009; Coe et.al., 2014), and professional identity (Oeberst & Imhoff, 2023; Askew et.al., 1997; Day & Gu, 2010). Evaluation is linked to professional development as the tool to identify effective and developmental areas across classrooms, schools, and multi-academy trusts (MATs). Professional knowledge is relevant in understanding effective evaluation processes and mathematics teaching. Professional identity acts as a lens to

understand individual perceptions and experiences as both the evaluator and the evaluated.

Research questions directly linked to the conceptual framework were devised to give clarity and structure to the study. These were simply, what are the perceptions and experiences of senior leaders, subject leaders and teachers of the evaluation of primary mathematics teaching in relation to professional development, professional knowledge and professional identity? Insight gained from exploring these questions was synthesised, discussed and analysed in order to gain deeper understanding of the four articulated problems.

Methodology

With its intention to explore the subjective perspectives and experiences of individual cases, this study sits within an interpretivist research paradigm and assumes a relative ontological and subjectivist epistemological position. The research design explored bounded case studies and used ‘fortune lines’ (Hall & Wall, 2019) and audio recorded semi-structured interviews as methods of narrative enquiry alongside documentary evidence to provide a more formalised framework against which to compare the accounts of each participant. Fifteen participants from five schools were recruited and equally represented the roles of senior leaders (SLs), mathematics subject leaders (MSLs) and class teachers (CTs), and all schools were members of the same MAT. Data were analysed using reflexive thematic analysis (Braun & Clarke, 2022) and according to the prior themes identified as key to the study through a review of relevant literature in relation to the conceptual framework.

Findings, Analysis and Discussion

The perceptions and experiences of senior leaders, subject leaders and teachers of the evaluation of primary mathematics teaching in relation to Professional Development

Whilst the importance of mixed data collection methods and triangulation of data was clear, there was a lack of clarity around the purpose of evaluations as formative or summative. The perceived credibility of the evaluator varied across role and school and was influenced by perceptions of both the knowledge held by the evaluator in terms of mathematics teaching, as well as the level of scrutiny that the methods and tools of data collection and interpretation were subjected to. Indeed, the credibility and validity of the tools and processes of evaluation were largely not considered with any coherence across groups. There was evidence that the power and autonomy over the evaluative process was largely held by SLs, with a particularly complicated role for MSLs in both a strategic and supportive position, and limited evidence of the evaluated being empowered to select the foci or methods for the evaluation of their practice. External evaluators were also perceived to hold a certain degree of power as the consequences of their validation influenced the perceived efficacy of teaching and informed targets or areas for development, although these were not always considered credible due to a lack of contextual knowledge.

The perceptions and experiences of senior leaders, subject leaders and teachers of the evaluation of primary mathematics teaching in relation to Professional Knowledge

In relation to professional knowledge, whilst there was a coherence in participants judgement of ‘effective/successful’ teaching regarding progress and attainment related outcomes, there was a wide variation in knowledge of what ‘good’ mathematics teaching is regarding the wider body of collective knowledge in the field. Depth and breadth of understanding of this was reliant on the personal and professional opportunities participants had had to develop such knowledge. As such, there was significant evidence of the influence of the knowledge prioritised by the National Curriculum and related tests, the growing prevalence of a ‘mastery’ pedagogy and approaches to the structure and sequencing of individual and series of lessons espoused by related schemes of work. There was largely an absence of a conceptualisation of mathematics that goes beyond these observable proxies of effective teaching. Therefore the shared criteria for evaluating the quality of teaching seemed to measure the implementation of these, rather than probing teaching and learning for deeper understanding. There was also lack of clarity, and in places coherence, between the curriculum of each school and the evaluation criteria, indeed no school used a subject-specific rubric as part of its formalised evaluation documentation. The foci for evaluations were selected from a range of generic and subject-specific criteria. Knowledge of robust and credible evaluation processes tended to mirror those employed by Ofsted and there was a lack of evidence that participants had knowledge of a wider evidence base to inform the valid and reliable collection and interpretation of data.

The perceptions and experiences of senior leaders, subject leaders and teachers of the evaluation of primary mathematics teaching in relation to Professional Identity

The personal relationship participants had with mathematics influenced and informed their professional identity in terms of their perceived self-efficacy, although there was evidence that many conceptualise ‘being good at maths’ as a separate and not necessarily precursive condition of ‘being good at teaching maths’. Perceptions of teaching efficacy and therefore the extent to which participants identified as a ‘good’ teacher of mathematics were influenced by evaluation experiences in both validating and invalidating ways depending on the level of congruence between the evaluated and evaluator’s beliefs about mathematics itself, and how it is best learned and taught. There was an acknowledgement in senior leaders of a tendency to evaluate the quality of teaching as consistent with their own beliefs and experiences.

Tensions between the multiple identities held by all participants were evident. Teachers with experience as evaluators, particularly those who had held leadership positions previously in their careers, found that the minimisation of this experience when being evaluated was challenging. Mathematics subject leaders found balancing their teacher and leader identities complex, particularly when in the role of evaluator with the higher levels of subject-specific knowledge but lower levels of autonomy over the foci and criteria of evaluations. Senior leaders for whom recent experience in teaching mathematics was lower expressed a lack of confidence in their judgements of the quality of mathematics teaching and a tension between this and the high level of accountability they felt.

The perceptions of autonomy and its relationship to identifying as a professional were largely related to being earned as a result of demonstrating compliance with the expectations of leadership and therefore trusted, rather than related to acknowledgement of experience. Expectations of compliance to schemes of work was viewed as restrictive to autonomy by more experienced teachers but necessary for the development of a coherent and consistent curriculum by middle and senior leaders, resulting in tensions around potential de-professionalisation in their use.

Implications and Recommendations

The following summaries outline the implications of the study as relevant to each of the aforementioned problems.

Problem 1: A potential lack of shared understanding of the varied purposes and consequences of evaluations

There was a predominant understanding that 'effective' or 'successful' teaching was aligned primarily with practices that produced desired learning outcomes as defined and limited by existing measurement processes (e.g., test scores). Consequently, evaluation processes focused heavily on identifying and replicating these 'effective' practices against a checklist, with less consideration given to defining and encouraging 'good' practice based on the broader body of mathematics education knowledge. This outcome-focused approach risks reducing evaluation to a tool for measuring and developing teaching against proxies assumed to guarantee results.

Problem 2: The impact of significant changes in mathematics curriculum, assessment and pedagogy on the evaluation process

The evaluation checklist of features was heavily influenced by the National Curriculum, end-of-key-stage tests, and the 'mastery' approach. While 'mastery' principles are often aligned with credible research, their practical implementation risks reducing the complexity of mathematics to a linear curriculum and a restrictively designed sequence of lessons. For evaluators and evaluands, this simplified delivery model can become a proxy for mathematical learning, influencing a surface-level understanding of teaching quality and overlooking deeper aspects of mathematical pedagogy.

Problem 3: Under-representation of key stakeholder voices in research and lack of clarity around the relationship between their roles

The perceptions and experiences of all individuals were highly subjective, leading to inherent tensions and contradictions. Achieving full consensus is unlikely, but greater coherence requires reflective dialogue that acknowledges the varying contributions of each stakeholder. Discussions must move beyond simply identifying the 'effective' to address what constitutes 'good' mathematics teaching and how to encourage it, considering purposes that transcend test-related outcomes and align with stated school visions.

Problem 4: The risk of de-professionalisation of teachers and leaders through the evaluation process

Teachers, MSLs, and SLs linked their professional identity to the extent they felt autonomous to make decisions consistent with their beliefs about effective mathematics teaching. However, a perceived need for consistency of practice created tension with this autonomy. Autonomy was often awarded not based on experience, but on demonstrated compliance with the delivery of specific proxies for learning. This focus on compliance and measurable outcomes limits the field to the 'effective' and risks reducing the professional trust afforded to many participants.

Model for teacher evaluation in primary mathematics teaching

What follows (Fig.1) is a recommended model for developing fair, useful and coherent evaluations (Nevo, 2006, 2015) of primary mathematics teaching which draws on Wenger's (1999) framework for community practice.

Mutual Engagement	Joint Enterprise	A shared repertoire of tools
Teachers, MSLs and SLs are equally included in discussions and decisions about 'what matters' in mathematics teaching evaluations, in terms of both 'good' and 'effective/successful' practice	There is mutual agreement on where, when and how evaluations take place, and on the foci for evaluation. Precedence is given to their relevance to learners' needs.	Use of a rubric closely aligned to the ultimate goal of high quality mathematical teaching.
Diversity, complexity and uncertainty are accepted as essential components of participation in this mutual endeavour.	The influence of external mandates is mediated through the community's shared negotiation of meaning. There is reduced reliance on proxies of effective mathematics teaching.	Use of appropriate data collection tools given the specific purpose of the evaluation. Precedence is given too tools that prioritise discussion, collaboration and space for pedagogical reasoning.
	There is mutual accountability as all involved are primarily accountable to developing fair, coherent and useful evaluations of teaching.	Evaluator development to interpret evaluation data and in increase awareness of and strategies to mitigate against subjective bias.

Figure 1: Model for the evaluation of primary mathematics teaching

Initially, all those involved in such evaluations will engage in reflective dialogue about the components relating to mutual engagement and mutual agreement. Subsequently, professional learning activities and opportunities to explore each of the components relating to the shared repertoire of tools would be provided, coupled with supported access to pertinent and credible sources of wider professional knowledge. Ultimately, such a model for practice has the potential to support teachers, mathematics subject leaders and senior leaders to move beyond a checklist when evaluating primary mathematics teaching and into richer professional communities of practice that acknowledge and respect the individual and collective wisdom of those involved contributing to the professional empowerment of all.

References

- Askew, M., Rhodes, V., Brown, M., Wiliam, D., & Johnson, D. &. (1997). *Effective teachers of numeracy: final report* (Feb 1997). King's College London.
- Braun, V., & Clarke, V. (2022). *Thematic analysis: a practical guide*. SAGE.
<http://prism.librarymanagementcloud.co.uk/chi-ac/items/532779>
- Coe, R., Aloisi, C., Higgins, S., & Major, L. E. (2014). *What makes great teaching? Review of the underpinning research*. October, 57.
<http://www.suttontrust.com/researcharchive/great-teaching/>
- Day, C., & Gu, Q. (2010). *The New Lives of Teachers* (1st ed). Routledge.
- Hall, E., & Wall, K. (2019). *Research methods for understanding professional learning*. Bloomsbury.
- Knowles, M. S., Holton, E. F., Swanson, R. A., & Robinson, P. A. (2020). *The adult learner: the definitive classic in adult education and human resource development* (9th ed.). London: Routledge.
- Levin-Rozalis, M., Rosenstein, B., & Cousins, J. B. (2015). A Precarious Balance: Educational Evaluation Capacity Building in a Globalized Society. In K. E. Ryan & J. B. Cousins (Eds.), *The Sage International Handbook of Educational Evaluation* (pp. 191–212).
- Nevo, D. (2006). Evaluation in Education. In I. F. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The Sage Handbook of Evaluation* (pp. 441–460). SAGE.
- Nevo, D. (2015). Accountability and Capacity Building: Can They Live Together? . In K. E. Ryan & J. B. Cousins (Eds.), *The Sage International Handbook of Educational Evaluation* (pp. 291–304).
- Oeberst, A., & Imhoff, R. (2023). *Toward Parsimony in Bias Research: A Proposed Common Framework of Belief-Consistent Information Processing for a Set of Biases*. *Perspectives on Psychological Science*, 174569162211481.
<https://doi.org/10.1177/17456916221148147>
- Rowland, T., Turner, F., Thwaites, A., & Huckstep, P. (2009). *Developing Primary Mathematics Teaching: Reflecting on Practice with the Knowledge Quartet*. SAGE Publications Ltd.
- Wenger, E. (1999). *Communities of practice: learning, meaning, and identity*. Cambridge University Press. <http://prism.librarymanagementcloud.co.uk/chi-ac/items/398002>