Reclaiming accountability through collaborative curriculum enquiry: New directions in teacher evaluation

Moira Hulme | Abigail Comber | Eli Jones | Julian Grant | John Baumber

Abstract
Teacher evaluation and teachers' professional learning are too often confined to separate areas of research and professional practice. Rather than approach evaluation and enquiry as distinct or irreconcilable, this paper applies the ideas of Stenhouse to explore new possibilities for the reappropriation of mandated appraisal in ways that support teachers' professional growth. Illustrative case studies of laboratory schools in the United States and England are used to examine the interaction of local and lateral forms of professional accountability with external and hierarchical regulatory frameworks. The article reports the design and enactment of change in two schools (a US kindergarten through twelfth grade school and a UK high school) connected through the International Association of Laboratory Schools (IALS) that purposively redesigned appraisal over a three-year period to build capacity for collaborative curriculum enquiry. Attention is afforded to the space for manoeuvre between advisory and mandatory guidance, and the challenges to relational trust and collective responsibility posed by performance-based accountability systems. The findings provide new insights into how teacher-led collaborative enquiry (curricular co-design) can address the unintended consequences of test-based accountability and rubrics-based observation as principal drivers of educational improvement.
INTRODUCTION

This article aims to reconnect areas of policy that are often treated as separate fields of enquiry. Drawing on Stenhouse's emphasis on curriculum as enquiry, we challenge the separation of curriculum re-design from strategies for professional growth, and the dislocation of teacher learning from evaluation of teachers and teaching. We argue that a commitment to student flourishing requires curriculum re-design that in turn requires recalibration of extant systems for appraising teacher performance. Previous studies of the formative and summative purposes of teacher evaluation tend to address only one goal (Tuytens et al., 2020). Reconciling accountability and improvement requires consideration of what educators should be accountable for, when and by whom, and which strategies are likely to support improvement. Different ways of conceptualising accountability and improvement reflect alternative models of professionalism in teaching. While the context of schooling has changed since Stenhouse championed school-based curriculum development, the imperative of collaborative deliberation on educational goals and purposes remains constant. We contend that current performative approaches to teacher evaluation derive from a deficit-based and overly simplistic model of teacher change and that it is timely to revisit alternative articulations of ‘experimental education’ premised on exploration rather than prescription.

There are strong individualising currents in teacher reform. Performance-based accountability measures have been introduced in advanced economies, notably the United States, the United Kingdom and more recently Australia (Clinton & Dawson, 2018; Cochran-Smith et al., 2018; Lingard et al., 2013; Rose et al., 2020). Policy attention has shifted from gate-keeping measures to protect a supply of ‘highly qualified’ teachers to performance management to retain and incentivise ‘highly effective’ teachers using the tools of rubrics-based principal observation and student growth scores (Martinez et al., 2016, p. 15). The rise of a ‘global testing culture’ (Smith, 2016) and the datafication of teachers’ work is indicative of declining trust in the competence of the teaching profession. Data profiles are increasingly used to support judgements about individual teacher ‘quality’ (Lewis & Holloway, 2019). Online data dashboards and digital communication platforms have introduced new forms of pervasive ‘dataveillance’ (Holloway, 2021, p. 39). Test scores are used as proxy quality measures in the ranking of teacher performance.

The use of standardised test measures in teacher evaluation proceeds from the assumption that student cohort performance can be correlated neatly with teacher competence. The attribution of student progress to individual practitioners disregards the influence of factors external to the school (socio-economic status, locational challenges, family health and wellbeing) and a range of within-school influences (class size, teacher turnover, professional learning opportunities, student prior learning experience, school budgets and resource allocation) (Berliner, 2018). Darling-Hammond et al. (2012) contend such attribution, ‘assumes that student learning is measured well by a given test, is influenced by the teacher alone, and is independent from the growth of classmates and other aspects of the classroom context’ (p. 8). School effects and locational challenges are silent partners in appraisal of individual performance.

Irrespective of construct validity, there are dysfunctional effects of performance measurement systems in public education. High stakes (i.e., consequential) metrics-oriented teacher evaluation has produced distortion or ‘gaming’ to obtain rewards or avoid sanctions. Testing drives effort substitution as the performance being measured attracts greater attention than...
unmeasured dimensions and, at worst, may incentivise individual and organisation-level cheating (Hibel & Penn, 2020; Taylor, 2021). Accountability systems promote ‘educational triage’ (Gillborn & Youdell, 2000) directing attention to those students who contribute most to changes at target thresholds. Moreover, pressure to improve measured outcomes may foster the constitution of ‘problem’ pupil groups and fuel an ‘intervention culture’ (Bradbury et al., 2021). Low-performing and/or vulnerable students may be excluded from higher stakes assessed activities (Hofflinger & von Hippel, 2020) and special education identification rates manipulated down to avoid external intervention (DeMatthews & Serafini, 2021). Examination-based success measures narrow curriculum choices in ways that are socially patterned and have ratcheted up pressure reducing the senior years of high school to a ‘two-term dash’ (Shapira et al., 2023, p. 43). Subsequent reductions in teacher job quality and job satisfaction have significant non-pecuniary costs that include student and teacher ‘emotional ill-being’ (Hargreaves, 2020 p. 393) and teacher turnover (Perryman & Calvert, 2020; Rubin et al., 2023).

In addition to test-based accountability, the outcomes of classroom observation are important sources of teacher evaluation data. Principal observation is a ‘ubiquitous’ but little understood practice (Lewis et al., 2022, p. 361). As Hazi (2022) notes, ‘both improvement and accountability are necessary functions for schools but can be difficult to accomplish in one practice of one person’ (p. 812). Principal preparation rarely provides substantive training in observational practice (Flores & Derrington, 2017). School leaders have limited time (and skills, sometimes) to provide high quality feedback to support sustained improvement action (Kraft & Christian, 2022). Where principal training is provided, gains in accuracy may not be matched by teachers’ perception of the usefulness of feedback (Goldring et al., 2020; Hermann et al., 2019). The administrative demands of school leadership mean that principals spend little time in their working week in classrooms. The implementation of comprehensive teacher evaluation schemes is particularly challenging in large high schools with diverse subject specialisms. Where principals do not position themselves as pedagogical leaders, evaluation practice can be reduced to ‘administrative technicalities’ (Lillejord & Borte, 2020, p. 284). Legitimate concerns around interrater reliability have encouraged the use of rubrics. However, an impulse to standardise and control disenfranchises educators if authoritative external schema displaces professional knowledge and deters peer dialogue. Holloway (2021) cautions that, ‘rubric-based observations are standardising our view of good teaching … having any single view of good teaching can be dangerous for enabling pluralism within schools’ (p. 34).

The performativity practices described above undermine teacher expertise, close down spaces for local deliberation and shift the locus of control away from the school (Gore et al., 2023; Holloway & Brass, 2018). The rhetorical positioning of twenty-first century teachers as ‘changemakers’, ‘curriculum makers’, and agentic ‘activist professionals’ (Burridge & Buchanan, 2022; Sachs, 2000) stands in tension with successive waves of policy that reduce teachers to implementers of changes pre-determined elsewhere (Elliott, 2001; Wrigley, 2018). Recent policies for school education in England and the United States have been subject to powerful decentralising and re-centralising tendencies that frame possibilities for local curriculum leadership. The disassembling of public education has produced a patchwork of free schools, multi-academy chains and charters. Self-managing schools purchase outsourced curricula, choose between external providers of teacher education, consultancy and testing services, and maintain a constant state of inspection readiness. Promising pedagogical practices are communicated through effect size lists that Biesta (2007) argues ‘not only restricts the scope of decision making to questions about effectivity and effectiveness, but that also restricts the opportunities for participation in educational decision making’ (p. 1).

This paper responds to the diminution of educator authority through market emulation models and regulatory frameworks that displace localised understanding of the needs of
school communities. It does this through reframing the problem of ‘teacher quality’ and its adjunct ‘teacher evaluation’ (Gore et al., 2023). Rather than approach evaluation and enquiry as irreconcilable we return to the ideas of Lawrence Stenhouse on ‘curriculum as enquiry’ and ‘research-based teaching’ to explore possibilities for reappropriation that challenge the ‘managerial drift’ in school leadership (Lillejord & Borte, 2020, p. 275).

The article is structured in four sections. First, we briefly explain how the ideas of Stenhouse inform our approach to collaborative enquiry in education work. Second, we outline how we enquired together as practitioner researchers to explore spaces for agency within performance-based accountability systems. Third, we examine the revisions that were made to the teacher evaluation process in two laboratory schools that sought to purposefully align curriculum development and appraisal for teacher growth. The final section revisits the literature on teacher evaluation and identifies implications for policy and practice.

**LAWRENCE STENHOUSE AND PROFESSIONAL ACCOUNTABILITY**

Stenhouse challenged educators to adapt and personalise rather than transfer and implement lessons from quasi-experimental designs derived from a ‘psycho-statistical paradigm’ (cited by Ruddock & Hopkins, 1985, p. 20). In the work of Stenhouse we find an early critique of prospects for universal pedagogy and the application of context-independent recipe knowledge of ‘what works.’ The model of extended professionalism expounded by Stenhouse positions teacher research as a professional responsibility and entitlement, in contrast with reductionist views of the teacher as technician charged with high fidelity curriculum ‘delivery’. Through an investigative process teachers are exhorted to seek deeper understanding rather than restrict professional enquiry to ‘how-to-do-it questions’ that produce ‘prescriptive answers’ (Stenhouse, 1975, p. 191). For Stenhouse, teacher learning was integral to improvement.

We know enough now to shun the offer of ready solutions. Curriculum research must be concerned with the painstaking examination of possibilities and problems. Evaluation should, as it were, lead development and be integrated with it. Then the conceptual distinction between development and evaluation is destroyed and the two merge as research. Curriculum research must itself be illuminative rather than recommendatory. (Stenhouse, 1975, p. 122)

Teacher collaboration and co-enquiry are at the centre of the model of democratic, extended professionalism advanced in the work of Stenhouse. His ideas on collective responsibility and shared enquiry are prescient in efforts to redefine successful school leadership in terms of achieving influence rather than simply securing compliance. His ideas challenge management models premised on solo leadership and recognise the value of aggregate teacher leadership. The next section explains our approach to collaborative enquiry in examining the strategies deployed by two schools to promote professional accountability and build collaborative capacity through curriculum making. Working against the grain, these schools sought to initiate and cultivate enquiry-led forms of professional development in a localised strategy that aimed to reduce perverse incentives towards competitive individualism.

**ENQUIRING TOGETHER: METHODS AND SOURCES**

To explore the intersection of curriculum enquiry and teacher development in an era of accountability we draw on the experiences of a purposive sample of two schools that took
professional growth as a starting point for redesigning teacher evaluation. Both schools are located in national contexts marked by decentralisation of the education state – the United States and England – where cultures of testing and evaluation are deeply rooted. In this sense, they represent atypical but critical cases. Within a whole school strategy, curriculum enquiry was introduced over three calendar years (2021–2023) to counter over-reliance on external rubrics-based assessments of teacher effectiveness and in acknowledgement of the complexity of teacher growth. While compliant with the mandatory state/national teacher evaluation requirements, school-level processes were purposively re-oriented to support teacher collaboration, build curricular capacity and promote cumulative knowledge production. This qualitative retrospective case study draws on first-person accounts and archival sources to articulate why these developments were undertaken, the levers available to school leaders, and draws on participant experience to provide insights into the intended transition from individual to collective accountability.

Data sources include revised curriculum frameworks co-designed to support student agency in learning, professional learning policies, resources and records of teacher enquiry groups, and written feedback generated through teachers' participation at school level. Teacher voice was an important feature of the re-design process. Every member of the teaching staff employed on full-time permanent contracts participated in the review of the teacher evaluation process at their school. Opportunities were provided for collective and individual contributions to the re-design process both within the process and outside/about the process e.g., verbal feedback in small group team meetings and whole school events, written feedback in teacher reflections (online proforma) and enquiry reports.

A secure digital archive was created containing the corpus of relevant textual files from structured engagements in both school settings (Table 1). The archive allowed the re-assembly of fine grained, situated local details for critical interrogation and reflection. In approaching documents, care was taken to consider congruence between the circumstances of creation and re-use (appraising data fitness and data quality), and risks of selection bias (Kutsyuruba, 2023). Written sources do not capture the visual cues of direct observation or the subtleties of speech interaction (Poland & Pederson, 1998). However, the high rates of

| Table 1 | Textual data sources 2021–23. |
|---|---|---|
| **School 1, US context** | **School 2, UK context** |
| Revised curriculum frameworks | Impact programme: trans curricular middle school programme (6th–8th Grade, 11–14 years) | Student Agency in Learning (SAIL) programme (Key Stage 3, 12–14 years) |
| Materials from teacher enquiry groups and professional learning sessions | Learning Communities in cognate clusters – 6 sessions per annum – involving whole staff | Cross-Faculty Enquiry Groups plus coaching – 9 sessions per annum – involving the whole staff |
| Teacher feedback/ reflections on revised teacher evaluation process (online proforma) | 2022/23, 46 teachers (100% teachers) | 2022/23, 20 teachers from total teacher roll of 48 (response rate 42%) |
teacher participation in producing reflective accounts and process of member checking lend credibility to the study's empirical claims.

Qualitative data analysis integrated deductive (theory driven) and inductive (data driven) coding (Nowell et al., 2017) and was conducted iteratively in three interrelated stages. The first stage involved careful reading and re-reading by Author 1 of relevant school policies, materials and reports. Document analysis was conducted to identify the components (content and context) and timeline for incremental school-level change in each organisational setting. Next, drawing on the literature on teacher collaboration and an eco-relational understanding of change (Daly et al., 2020) thematic codes were tested against leaders' artefact-based reflections and teachers' written reflections of their experiences of change within and between the two school settings. In identifying patterns of meaning across the data, attention was afforded to the preconditions, drivers and challenges of desired change and participant experiences of the process. As the research is motivated by interest in the scope for leader agency the analysis attended to the space for manoeuvre available to school leaders as they navigated fidelity with external regulatory regimes and sought to promote the collaborative design and collaborative enquiry on curriculum alternatives. Space for manoeuvre is defined by Priestley et al. (2012) as, ‘personal capacity to act, combined with the contingencies of the environment within which such action occurs’ (p. 196). Analysis was an iterative process supported by email and online video conference communication across the team between April and September 2023 to interrogate interpretations. The inclusion of multiple perspectives and data sources and the involvement of practitioner researchers looking across the two settings adds to the ‘trustworthiness’ of the findings (Bryman et al., 2008). Collaboration between university and school-based researchers supported critical self-reflection, reflexivity and discussion of positionality (Pillow, 2003). Investigator triangulation combined the externality of university-located researchers with the proximity of school-based research leaders. Sense making was checked across team members for consistency and accuracy.

The research was informed by the British Educational Research Association's Guidelines for Educational Research and the Code of Ethics of the American Educational Research Association (AERA, 2011; BERA, 2018). Approval was sought through the University Ethics Committee, approval certificate number 00156. Consent procedures acknowledged an intention to publish, the limits of anonymity when identifying the participating schools, and the right not to participate or withdraw contributions from the study at any time without penalty. The school processes reported are subject to ongoing review and adaptation and will continue to evolve through school-based deliberative processes.

The schools

Burris Laboratory School was established in 1929 as part of Teachers College, Ball State University, and Muncie Community Schools, Indiana. Burris was named after Dr. Benjamin Burris, Dean of Ball State's Teachers College when the school was built. The school is a highly rated coeducational public school that caters from kindergarten to grade 12 (students aged 5 to 17/18 years) with a student-teacher ratio of 14 to 1. In 2023/24 the school had 50 teachers and 672 students. The Laboratory School acts as model school offering practicum experiences for pre-service teachers and has the midwestern state of Indiana as its enrolment district. Students are admitted by a lottery system. Burris does not have a school board. The Ball State University Board of Trustees is the ultimate authority for school policies.

Shevington High School is a coeducational community secondary school (students aged 11–16 years) in the northwest of England. The school was rated Good by the national school inspectorate, the Office for Standards in Education (Ofsted) in its last four consecutive inspections (the most recent being March 2022). In 2023/24, the school had 844 students and
RECLAIMING ACCOUNTABILITY THROUGH COLLABORATIVE CURRICULUM ENQUIRY

48 (full time equivalent) teachers, a student-teacher ratio of 17 to 1. The school is a local authority, maintained school that has chosen not to partner with a federation or multi-academy trust. The school is involved in teacher education in partnership with Kingsbridge Teacher Training, a School-Centred Initial Teacher Training (SCITT) provider. As a community high school, the Governing Body are drawn from the local community – parents, teachers, local authority, community groups.

Both schools are members of the International Association of Laboratory Schools (IALS), an elective (subscription based) international association of schools committed to professional enquiry and educational experimentation (www.ials-laboratoryschools.org). IALS adopts an expansive definition of laboratory schools that includes ‘campus-based schools, and others with diverse university affiliations, such as charter schools, professional development schools, child study institutes, research and development schools’ (n.pag).

RECALIBRATING EDUCATIONAL EVALUATION IN TWO LABORATORY SCHOOLS

In this section, we draw on documentary sources and the words of participants to convey how each school attempted to use curriculum enquiry as a lever to foster teacher collaboration. We consider the constraining and enabling contextual factors that influenced how far teacher evaluation shifted beyond episodic assessment of individual performance towards collective responsibility and sustained participation in joint work.

From performance rating to learning communities

School teachers in Indiana are accustomed to looking at growth scores for their students. Individual growth model measures are available for students and teachers in English Language Arts and Mathematics in grades 4–8 (students aged 9–13 years). From 2012 the Senate Enrollment Act (Public Law [PL] 90) introduced a teacher evaluation system that employed Value Added Measures and rubric-based assessment within a merit pay structure. Students' growth scores were used within annual evaluations to help situate teachers in one of four rating categories: highly effective, effective, improvement necessary and ineffective (Indiana Department of Education, 2020). Performance ratings were determined by a trained evaluator in locally selected competencies correlated with student outcomes. Teachers rated less than effective faced a salary freeze and potential dismissal (Harvey et al., 2019). From 2020, with no test scores for that year due to the pandemic, Indiana ended the requirement of test scores being tied to evaluation and teacher effectiveness.

At Burris laboratory school exposure to increased individual accountability was used as a stimulus for collective deliberation on what constitutes student success and how this might be advanced. When the mandate was lifted, teachers elected to retain elements of student growth to gauge their effectiveness in the classroom. However, learning progression data was now used for developmental purposes within professional deliberation in school, rather than ‘weaponised’ for external accountability. When setting Student Learning Objectives, educators need to have a good understanding of students' starting points and of assessment practice. Skill is needed in appraising learners' starting points and selecting appropriate teacher-created and schoolwide assessments and tiered targeted goals to establish progress. Sources of conventional evidence include beginning-of-course diagnostic tests or performance tasks, results from prior year teacher-, school-generated and state tests that assess knowledge and skills that are pre-requisites to the current subject/grade (Evaluator and Teacher Handbook v3.1, 2020, p. 21). The purposive use of diverse and multiple sources
Curriculum innovation followed teacher-led deliberation on local assessment practices and curriculum intentions. The school exercised curricular flexibility to create a middle school initiative (students aged 11–14) called Impact to run alongside state required core content. Impact is a trans curricular programme (addressing multiple subject disciplines through a wide variety of content and units) where students are blended in classrooms from sixth to eighth grade. From 2021–2022, 12 were teachers involved in the programme; four at each grade level (6th–8th). Curriculum re-design was intended to promote intrinsic motivation through personalised inquiry-led learning. The curriculum is organised around five pillars: individuality, metacognition, purpose, community and transfer. Using the strapline of ‘ungrading though ZPDs’ (zone of proximal development) teachers are encouraged to engage in curriculum making that advances personalised learning, individualised instruction, student-led progress monitoring and real time feedback (Kohn & Blum, 2020). To foster collaboration and deepen professional relationships, Impact teachers were given 1 day per week (to support peer observation, co-teaching and joint work), which includes three scheduled hours of shared time for focused collaboration.

From 2021, the teacher evaluation process was progressively refined to incorporate and value an explicit enquiry stance that more closely aligns with local curricular expectations. In the first iteration, teachers self-organised into learning communities (LCs) and participated in self-reflection on growth towards teacher-created goals. The following year, teachers were grouped in learning communities based on similar goals to strengthen support for and coherence across teacher inquiries. Teachers selected two goals, one identified with administrators and another self-identified and posed as a question. For example, a teacher who wishes to work on developing student engagement might ask, How does student self-assessment affect student engagement in classroom activities? Admin-created goals and enquiry-focused goals carry equal weighting. In the third year, teachers selected two goals, and joined learning communities based on the admin-determined growth goal to balance both individual and school-identified priorities. Members of the LCs visit each other’s classrooms for peer observation and meet six times a year. Reflections are formally collated twice per annum and include a range of teacher curated artefacts that best evidence student impact and professional growth. A digital secure drive is used to save and share multimodal files between community members. Based on teacher feedback, the model was further strengthened through the provision of scheduled time for collaborative meetings and more specific information to guide reflection.

In spring 2023, teachers’ written reflections (online proforma) on their experience of the revised approach to evaluation suggest progress towards its intended aims. Over 3 years, the initiative enabled a move from over-reliance on quantitative assessment using test scores as a proxy measure of teacher quality, which was felt to incentivise a narrow focus on teaching to test. An enquiry orientation to evaluation encouraged exploration rather than standardisation of teaching approach.

In a previous version of the evaluation programme, the emphasis was almost solely quantitative. I needed to select a class to achieve a goal, give a pre- and post-assessment, and assess whether students made that goal solely on the data. I felt that my growth as a teacher boiled down to if enough students answered a few questions correctly. With the new approach, I can focus on a much more holistic, qualitative goal. I have the freedom to try different methods that may or may not work. Instead of focusing on if students can answer questions on a test correctly, I can focus on how they are learning and their level of engagement. (Teacher, Middle and High School, Maths and Science)
By affording greater flexibility over identification of goals, many teachers felt more able to locate themselves within the process. This may support the construction of a more coherent teacher identity that does not rest on compliance with a perceived model of the ‘effective’ practitioner embodied in a Standards narrative developed outside and independent of the context of practice. Self-directed problem posing helped to shift the locus of control towards the teacher as enquirer.

I find that this model allows me to be me in the classroom. I’m not worried about fitting into a mould that ‘looks’ good on paper. I’m not reviewing the rubric and tailoring my instruction around it for the inevitable and unpredictable in class evaluation. (Teacher, High School, Arts and Social Studies)

At a pragmatic level, the careful articulation of a single personal goal was regarded as manageable and reduced a tendency among some practitioners to overcommit to improvement action across multiple targets set by others.

This new evaluation process has given me the freedom to focus on one personal goal at a time and master that skill or goal. Before, I felt overwhelmed trying to do too many things, and then I didn’t do anything to the best of my ability. (Teacher, Middle School, Technologies)

Teachers valued the opportunity to engage as an active participant in what was perceived to be a co-owned endeavour. Local learning communities opened spaces for experimentation and dialogue that were restricted in their previous experience of management-led modes of appraisal.

I appreciate that our evaluation is authentic. It is not the proverbial ‘dog and pony show’ of the past. I am able to be an active participant by collaborating with admin and colleagues. I love that I may choose the avenue by which to meet my goals and how that affords me opportunity for research and to seek out different professional growth avenues. My goals are personal and relevant. (Teacher, Elementary School)

Our format for teacher evaluation has allowed me to pursue a passion I have in [my subject] and use it to become more focused with my lessons. It has given me the platform to experiment and try new things without fear of being judged or seen as a failure. This has helped to put the focus on the process and not always the final project. I can reflect on my personal growth with intention. (Teacher, Elementary and Middle School, Creative Arts)

The iterative development of the evaluation model increased the legitimacy of the process. Teachers did not fear ‘retaliation’ when providing feedback to each other and school leaders, because the revisions to the school evaluation model started first with teachers, and teachers’ views and experiences shaped each iteration. Aligning curriculum enquiry with the teacher revised evaluation system and deepening engagement overtime, helped to foster trust. Reappropriating learning progression data for formative, developmental purposes and protecting space for teacher-created goals was a public expression of trust in teacher competence. Similarly, prioritising teacher voice in shaping the direction and pace of change contributed to the perception of openness and responsiveness. Attention was afforded to community building as a social practice through the provision of scheduled time for collaboration.
From appraisal to enquiry groups

At Shevington High School mandatory teacher evaluation was recalibrated to focus on professional growth using the strapline of ‘teacher appraisal though effective professional development’. In England, teacher performance is appraised using teacher-generated information, assessment data and observation using the Teachers’ Standards (DfE, 2011). Progression decisions are made by the Pay Review Committee of the school/MAT Governing Body and consequences for underperformance can include pausing pay progression, initiation of capability procedures and ultimately dismissal (DfE, 2019). From January 2022, the school sought to reconfigure experiences of appraisal to emphasise its key role in supporting teacher development, and to better align professional learning strategies with the school’s curriculum vision.

The transition was supported by work across the previous year (2021/22) using an external facilitator to introduce instructional rounds (City et al., 2009) and appreciative inquiry to build staff confidence, receptiveness and capacity to engage in collaborative dialogue. The emphasis on appreciative inquiry within internal accountability processes acknowledged the challenges in establishing trust within complex systems. Creating the conditions for agentive professional learning was aligned with a commitment to promote student agency in learning. Student agency is defined here as ‘the capacity to set a goal, reflect and act responsibly to effect change’ (OECD, 2019, p. 2). Shevington was among 20 schools in England that were part of a Student Agency in Learning Project (SAiL) (2017–2023) influenced by the Swedish Kunskapsskolan Education (KED) programme (www.kunskapsskolan.com/thekeprogram). A model of student coaching and personalisation was introduced through participation in this network. Timetable space was created for every student at Key Stage 3 (aged 12–14 years) to have 30-min ‘base group’ sessions 4 days a week. These sessions are used to set personally identified ‘learning missions’ and are intended to promote curiosity and independent working skills. In addition to a required core curriculum (i.e., English Language, English Literature, Mathematics, Science, Physical Education, Personal, Social and Health Education and Citizenship), from 2023/24 the school has prioritised a broad range of elective subjects at Key Stage 4 (aged 15–16 years) that includes Art, Drama, Music, Creative Media, Photography, Film Studies, and Mandarin. Curriculum breadth allows students to pursue two Arts subjects or two Languages.

The re-design process entailed integrating and reorienting required appraisal procedures within the activities of curriculum-focused Enquiry Groups. Each group is comprised of six teachers who are supported to set teacher goals and complete action research. The process is coordinated by a teacher recognised by the school community as a highly skilled practitioner who was given a Teaching and Learning Responsibility (TLR) salary increment. In addition, two early career teachers provide complementary support through focusing reviews of curriculum maps, schemes of work and enquiry plans, and by providing timetable support for peer observations in the spirit of appreciative inquiry. The co-leads are purposefully not members of the school senior management team so that the process would not be deemed as a top-down further quality check by management. The system is designed to encourage horizontal peer-to-peer open and authentic dialogue. Outcomes-based accountability within hierarchical line management structures was considered less conducive to the identification and exploration of areas for development that may be resistant to change. The goal is to embed a sustainable process of enquiry within the day-to-day life of the school community. Supported joint work in teacher identified priority areas was intended to enhance the strength and quality of relational relations between teachers.

Teachers are encouraged to identify challenging areas to subject to systematic enquiry: specifically, areas that are less amenable to conventional metrics such as fostering intrinsic motivation. Enquiry groups feature one teacher from each faculty, where possible. An
earlier system of triads was expanded to groups of six to ensure continuity in the event of staff absence. A 'live' digital CPD profile is presented as a learning journey and support is offered through structured milestones e.g., stages in the action research cycle (plan-act-observe-reflect) and through protected time at four coaching points across the year. A combination of internal (November and February) and external professional learning opportunities (May) is designed to be responsive to emergent teacher identified needs. The Performance Management Review is no longer a separate process but is integrated into the teachers' CPD profile, making collecting evidence throughout the year easier and avoiding duplication. Multiple sources of evidence are sought that extend beyond the conventional data dashboard (screening data, assessment and attendance data) to include records of student voice engagement activities, peer observation records, materials from targeted interventions and work samples. The importance of action research within teacher evaluation was strengthened in the wake of the pandemic when the Pay Review Committee focused on professional development more deeply than in previous years due the vagaries of available assessment data. The direct and continued engagement of Governors with teacher research gave credibility and kudos to the activity.

Written reflections on the process of change gathered through online proforma suggest teachers find the process 'more personal', 'more tuned to CPD' (Teacher, Maths and Science), and enabled them to have 'more ownership and direction over what I wanted to achieve over the year' (Teacher, Social Studies). Enquiry-led in-situ professional learning was deemed more relevant in addressing specific local concerns: 'The enquiry question prompted me to be more active with my CPD without having to have external training' (Teacher, Maths and Science). Structured support throughout the year to identify, interrogate and collate diverse sources of evidence was valued. The digital profile made it 'easier to collect other evidence' (Teacher, Maths and Science), and ‘evidence is on-going not just at the end of the cycle, not reliant on Year 11 data' (Teacher, Arts and Humanities). Opportunities for ‘cross departmental collaboration’ and ‘help from more experienced colleagues’ (Teachers Maths and Science) helped build a sense of community and collective responsibility.

We identified an area of research that interested us individually. The priority group connected to that area of interest encouraged further co-operation and provided valuable connection with others to work collaboratively to find solutions. (Teacher, Maths and Science)

Not all teachers embraced the opportunity to engage in enquiry and some felt more comfortable with management-directed activity. For a minority of teachers, the relationship between enquiry and the demands of day-to-day practice continued to be oppositional. Under time pressure and fatigued, it appeared for some that difficult choices were needed between focusing on senior year exam classes and committing fully to undertaking professional enquiry.

I have struggled a bit with the ‘Enquiry Group’ aspect. I struggled to decide on a topic and found the self-directed process took up a lot of time. I have not been able to do as much as I would have liked due to focusing on Year 11. I would rather there be a choice of set topics/courses’ to follow that are pre-set and can be worked through. (Teacher, Arts and Humanities)

Teachers at Shevington produce an annual research summary to share with their colleagues and their reviewer. From 2021–22, in July each year, teacher action research is presented at a whole school event in which emerging claims are subject to peer feedback, ways forward are identified, and research engagement is celebrated. Credible enquiry outcomes...
are used to support departmental plans and the school improvement plan. As the initiative develops the school is generating a resource of previous teacher enquiries, creating possibilities for multiple local case studies and cumulative knowledge building. A digital shared drive contains clusters of teacher enquiries organised in cognate themes. Leaders of professional learning in school can use the archive as a resource to support teacher induction and ongoing enquiries. This collective archive does not rest with individuals and key learning is not lost when teachers leave the school.

**DISCUSSION AND CONCLUSION**

The sections above outline the response of two laboratory schools operating in different national contexts to a marked performative turn in teacher evaluation. Consequential systems for teacher evaluation (for progression, compensation and tenure) have been widely adopted in many countries. However, despite policy advocacy there is limited evidence of positive impact on student outcomes or teacher learning (Berliner, 2018; Bleiberg et al., 2023). An international review of classroom observation systems by Martinez et al. (2016) concludes ‘There is little empirical evidence to suggest that large-scale standardized observation, and high-stakes teacher evaluation in general, constitutes a good vehicle for countries to exert improvements and achieve educational success’ (p. 27). Indeed, accountability systems that purport to ‘close the gap’ are associated with occupational stress and teacher attrition, particularly in socio-economically disadvantaged settings (Glassow, 2023).

This study was undertaken by practitioner researchers based in school and university settings to reflect on the opportunities and challenges of doing ‘appraisal’ differently. Powerful processes of responsibilisation construct modes of response that are both reactive – accounting for performance – and proactive – taking of responsibility for one’s own professional growth. Educators possess a capacity to respond that may be more managerial or democratic, more influenced by market-oriented or relational accountability. However, such well-rehearsed binary oppositions provide poor sources of support for practitioners charged with working creatively and ethically within regulatory frameworks. Exhortations to activism underestimate multiple constraining influences. This study provides new insight into how school communities can build and sustain space for agency as external scrutiny intensifies. Lewis et al.’s (2022) metaphor the ‘mantle of agency’ is helpful in re-positioning school leaders as ‘neither complicit nor disengaged, but rather able to act in accordance with their professional knowledge and judgment, with policy as cover’ (p. 362). The cases presented here show how school leaders donned the ‘mantle of agency’ to adapt evaluation systems to actively promote teacher learning. School leaders appraised the opportunity costs of peer collaboration and took advantage of windows of opportunity to realign mandated processes in accordance with desired education goals (e.g., the presentation of richer sources of evidence to the Pay Review Committee post-pandemic and refocusing data discussions within curriculum inquiry). Leaders need courage to differentiate between guidance and mandatory requirements, for example, in exercising discretionary powers to widen the curriculum offer, prioritise student agency in learning, and sustain inquiry through the senior years of high school.

Transitioning from evaluation systems informed by competitive individualism presents significant challenges. The moves described above were incremental, developed over several years and continue to be subject to iterative refinement. For example, peer review was first introduced at Shevington using appreciative inquiry in instructional rounds to foster a culture that might sustain higher levels of peer-to-peer open dialogue. The revised system for teacher appraisal developed from and strengthened ongoing curriculum redesign to protect and promote student agency in learning. Teachers were given opportunities to engage in
supported joint work. Opportunities for collaboration were by design and organic, formal and informal (e.g., shared social breaks such as fika Friday’s at Shevington and Teacher Tailgate socials at Burris). Structures were intentionally reconfigured to enable change without recourse to positional authority alone. This is consistent with emerging evidence of the positive impact of teacher-led evaluation on teacher motivation (Ford & Lavigne, 2023). Team-based collaborative professional learning opportunities are more effective in changing classroom practices (Kennedy, 2019; Lillejord & Borte, 2020). Gore et al. (2023) demonstrate the positive impact of reciprocal participation and collaborative engagement in extensive research on Quality Teaching Rounds. Widening the pool of observers (through supportive peer observation within collaborative enquiry) and the removal of responsibility for coordinating peer review from senior management reduces tensions that arise when school leaders are positioned as both supervisor (supporter) and evaluator (judge) (Lewis et al., 2022). The shift towards horizontal accountability within learning communities and enquiry groups helped to reduce the risk of hierarchical ‘judgementoring’ (Hobson & Malderez, 2013).

Not all teachers will embrace an inquiry stance and the early stages of change may involve a degree of teacher mobility. Numerous writers have cautioned of the risks of co-option and superficial accommodation variously described as ‘contrived collegiality’ (Hargreaves, 2019, p. 609), ‘cordial hypocrisy’ (Solomon & Flores, 2001, p. 13), ‘passive buy-in’, ‘fabricated cooperation’ and ‘collusion’ (Chapman, 2019, p. 7). Quality and depth of collaboration is influenced by the capacity and receptiveness of participants, previous experience of change initiatives, institutional characteristics and culture, the size and structure of collaborative arrangements, as well as the affordances of the external policy environment. In both schools, teachers were initially supported (with time and resources) to collectively re-engineer and evaluate the middle school/Key Stage 3 curriculum offer before extending opportunities for focused small-scale experimentation across the wider school community. Multiple accountabilities require teachers to move between ‘calculative’ modes of engagement (self-interest) and strategic investment of effort (e.g. the two-term dash in exam classes) and those based on ‘relational’ trust or community benefit (Fink, 2016). Professional bonds based on relational trust need to be nurtured constantly, and not just inscribed in local teacher development policy. As Kutsyuruba et al. (2016) note, ‘the dynamics of trust are expressed through changes in scale and intensity over the course of time and relationships, as expectations are either fulfilled or disappointed and as the nature of the interdependence between people changes’ (p. 345). To grow opportunities for teacher leadership continued vigilance is needed to sustain openness and guard against the creation of new enclosures through slippage of original goals, and the creeping enfolding of external authority over time and staff changes. If educators are to reclaim authority over their work this will require stable conditions that validate and support constant problem posing. This was initiated in these schools through the reappropriation of a high-profile mandated activity such as appraisal and its deliberate reworking in the service of professional enquiry over a three-year period. This entailed broadening the base for peer observation and providing opportunities for curriculum leadership.

Creating and sustaining conditions conducive to curiosity-driven professional learning is difficult within a prevailing school testing culture. It is a more complex task than telling teachers what to do through recipe knowledge of what works or the application of standardised observation rubrics of effective practice. Authority-based teaching as a form of answerism is a poor basis for continual professional growth. Stenhouse (1979) asked more of professionals when he insisted that ‘Research-based teaching is more demanding than teaching which offers instruction through a rhetoric of conclusions’ (reprinted in Ruddock & Hopkins, 1985, p. 113). In contrast to authority-based teaching, enquiry-based teaching places the teacher as a ‘seeker’, ‘questioner’ and ‘researcher’ (ibid. p. 119). Through the reappropriation of appraisal and advocacy of collaborative curriculum enquiry these two laboratory schools sought to move research to the forefront of teachers’ work as curriculum makers. By placing
enquiry at the centre of a reinvigorated model of professional accountability, evaluation activities are no longer individualised but a community effort, no longer private but public with the audience being other educators. In navigating accountability requirements, the concept of curriculum as enquiry provides a useful framework for those who seek to align systems for the evaluation of teachers’ practices and activities that promote professional growth.

FUNDING INFORMATION
This research received no specific funding or grants.

CONFLICT OF INTEREST STATEMENT
This study reports practitioner research, where three co-authors are employed by schools that are the focus of the study.

DATA AVAILABILITY STATEMENT
Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ETHICS STATEMENT
Approval was granted by the University of Bolton (employer of lead author at the time of the research), certificate no. 00156.

GEOLOCATION INFORMATION
England and USA.

ORCID
Moira Hulme https://orcid.org/0000-0002-4095-3880

REFERENCES


---

**How to cite this article:** Hulme, M., Comber, A., Jones, E., Grant, J., & Baumber, J. (2024). Reclaiming accountability through collaborative curriculum enquiry: New directions in teacher evaluation. *The Curriculum Journal, 00*, 1–17. https://doi.org/10.1002/curj.272