BMJ Open

A realist evaluation of UK medical education quality assurance

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-033614
Article Type:	Original research
Date Submitted by the Author:	14-Aug-2019
Complete List of Authors:	Crampton, Paul; Hull York Medical School, Health Professions Education Unit; UCL, Research Department for Medical Education Mehdizadeh, Leila; UCL, Research Department for Medical Education Page, Michael; QMUL, Institute for Health Sciences Education Knight, Laura; UCL, Research Department for Medical Education Griffin, Ann; University College London, Medical School
Keywords:	QUALITATIVE RESEARCH, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, MEDICAL EDUCATION & TRAINING

Title page

Title: A realist evaluation of UK medical education quality assurance

Authors: Paul E.S. Crampton[1,3] , Leila Mehdizadeh[1] , Michael Page[4] , Laura Knight[1] , Ann Griffin[2]

Address affiliations:

[1] Research Department of Medical Education, UCL Medical School, Royal Free Hospital, Room GF/664, London, NW3 2PF, UK

[2] Research Department of Medical Education, UCL Medical School, The Directorate, 74 Huntley Street, London, WC1E 6AU, UK

[3] Hull York Medical School, York University, John Hughlings Jackson Building, University Rd, Heslington, York YO10 5DD, UK

[4] Institute for Health Sciences Education, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, Room 3.15, 3rd Floor, Garrod Building, Turner Street, Whitechapel, London, E1 2AD UK

Corresponding author: Ann Griffin, Research Department of Medical Education, UCL Medical School, The Directorate, 74 Huntley Street, London, WC1E 6AU, UK

Word count: 3978

Key words (MESH): Quality assurance, health care; Qualitative research; Medical Education

ABSTRACT

Objectives: The aim of the study was to explore what components of the General Medical Council's (GMC) quality assurance framework work, for whom, in what circumstances and how?

Setting: UK undergraduate and postgraduate medical education and training.

Participants: We conducted interviews with a stratified sample of 36 individuals. This included those who had direct experiences, as well as those with external insights, representing local, national and international organisations within and outside medicine.

Intervention: The GMC quality assure education to protect patient and public safety utilising complex intervention components including meeting standards, institutional visits and monitoring performance. However, the context in which these are implemented matters. We undertook an innovative realist evaluation to test an initial programme theory. Data were analysed using framework analysis.

Results: Across components of the intervention, we identified key mechanisms including: transparent reporting to promote quality improvement; dialogic feedback; partnership working facilitating interactions between regulators and providers, and; role clarity in conducting proportionate interventions appropriate to risk. The GMC's framework was commended for being comprehensive and enabling a broad understanding of an organisation's performance. Unintended consequences included confusion over roles and boundaries in different contexts which often undermined effectiveness.

Conclusions: This realist evaluation substantiates the literature and reveals deeper understandings about quality assuring medical education. While standardised approaches are implemented, interventions need to be contextually proportionate. Routine communication is beneficial to verify data, share concerns and check-risk; however, ongoing partnership working can foster assurance. The study provides a modified programme theory to explicate how education providers and regulators can work more effectively together to uphold education quality, and ultimately protect public safety. The findings have influenced the GMC's approach to quality assurance which impacts on all medical students and doctors in training.

STRENGTHS AND LIMITATIONS

• Quality assurance of medical education remains an expensive process yet there is very limited research to understand its effectiveness

• This study, underpinned by a sound team-based reflexive analysis, is the first in-depth realist evaluation of quality assurance in the healthcare context which is responsible for all medical doctors and students in training

• We found that depending on context, the same interventions triggered a range of mechanisms leading to positive or negative outcomes

• The study is specific in its UK focus however we collected qualitative data from a large number of UK based and international expert stakeholders within and outside medicine, enhancing the transferability of our findings

• While standardised approaches are implemented, interventions need to be contextually proportionate. Routine communication is beneficial to verify data, share concerns and check-risk; however, ongoing partnership working can better foster assurance to protect patient safety

INTRODUCTION

Rationale

Healthcare regulators quality assure education and protect public safety utilising complex intervention components including setting standards, institutional visits and monitoring performance. However, the context in which these components are implemented matters.[1-3] Within undergraduate and postgraduate medical education, the taught curriculum integrates with workplace-based experiential learning. Consequently education environments range from structured classrooms in university contexts to clinical placements within shifting healthcare structures. Therefore, the challenge for regulators is to mediate the quality of education and training across these spaces in order to assure the public that education and training is safe.

In the UK, the General Medical Council (GMC) work closely with other organisations to secure its standards, using a three-tier model. The GMC (Tier 1, quality assurance), has an overarching responsibility to hold undergraduate and postgraduate training bodies to account for meeting standards. These bodies (Tier 2, quality management) organise, manage, commission, and sometimes deliver medical education. They also manage quality in local education providers (LEPs), where students and trainees are placed, such as trusts, health boards, general practices, and other clinical settings. The LEPs (Tier 3, quality control), have processes to ensure satisfactory clinical placements, and that their organisation provides an appropriate learning environment. Medical royal colleges work with the GMC to ensure their curricula and assessments are fit for purpose, inform specialty and postgraduate programme delivery, and have local systems to support training.

The GMC has a multifaceted intervention to examine the quality of medical education provision, known as the quality assurance framework (QAF, figure 1). The intervention includes components: setting standards, approving education settings, monitoring activities including self-assessment and enhanced monitoring, visits, sharing evidence with other regulators and identifying good practice.[4] The QAF operates across the three-tier model i.e. between the regulator (e.g. GMC), organisational provider bodies (e.g. medical schools) and local service delivery organisations (e.g. hospitals, general practice).

Figure 1: The intervention: quality assurance framework (QAF)

[Insert figure 1 here]

A range of approaches can be implemented to assure education quality, from heavily arbitrated measures to informal uncontrolled processes. Existing education Quality Assurance (QA) research is sparse and tends to come from the field of school-based and higher education.[5 6] Whilst exploring the mechanisms of action of school inspections,[5] a theory stated that regulatory activities associated with improvement include: setting standards; the provision of feedback; employing a system of sanctions and rewards; monitoring schools by the collection of information and public accountability. However, more research is needed to understand within the healthcare context how quality assurance can protect patients. Despite large amounts of resources dedicated to education QA, there remains a lack of clear evidence. QA takes place within varied and complex social environments. For this reason, the same intervention can impact on individuals, teams and organisations in different ways.[7 8] Although there are intended consequences explicit in the QAF design, the implicit underlying drivers of these are not clear.

Specific aim

The study aim was to explore what components of the GMC's quality assurance framework work, for whom, in what circumstances and how?

METHODOLOGY

Conceptual framework: realist evaluation

We chose realist methodology because of a focus on four theoretically constructed and inter-related questions: what works, for whom, in what circumstances, and how?[9-11] This results in generative causation, about how QAF components operate, offering an assessment of *whether* they work, as well as *why*. In the results, we explore the complex configuration links between contexts (where, when and with whom the activity takes place), components (different activities applied to assure quality), mechanisms (underlying processes for why the activity is/is not effective), and outcomes (intended and unintended consequences).[12] Our methods follow the RAMESES 2 (realist and meta-narrative evidence synthesis: evolving standards)[12] reporting guidelines (for full report see Griffin).[13]

Initial programme theory

We developed an initial programme theory (figure 2) based on existing literature, the GMC's approach to QA and research team insight (see reflexivity). We positioned the QAF as consisting of various components which we then explored to answer our study aim. Each component triggers multiple responses when applied in certain contexts with underpinning resources. Our programme theory postulated that within undergraduate, postgraduate and local education provider contexts, the QAF led to improved quality and protected the public through exploiting regulatory influence, guidance and supporting organisations, leading to compliance, resistance, relationships and empowerment.

Figure 2. Initial programme theory

[Insert figure 2 here]

Sampling and recruitment

Stratified purposive sampling was used to test our theory with stakeholders. We targeted those who were familiar with the QAF, labelled in the study as quality assurance partners (QAPs); as well as those with outsider perceptions of how the framework is positioned in society (e.g. other regulators) and broader regulation contexts (e.g. education). These spanned organisations both inside and outside healthcare, including internationally; collectively labelled as non-quality assurance partners (non-QAPs). See table 1 for the list of QAPs.

Our realist position acknowledged that stakeholders each had partial knowledge of the intervention, therefore to fully explore research questions we included policy makers, implementers and recipients.[14] Following email invitations, non-responders received two reminders. Aligned to realist evaluation,[15] participants were given the opportunity to review project materials prior to participation via a 15-minute informational video.

Data collection

To test our programme theory we undertook semi-structured interviews to explore underlying processes triggered by QA components (Supplementary files_Interview schedule v1.0). We posed a number of candidate theories e.g. to test theories on generalisability and utility[16] we asked: "Some hold the view that organisational self-assessment is not a reliable process, what do you think?", and "What would happen if the medical regulator (GMC) did no organisational visiting?" We designed questions with different foci appropriate to participants,[15] for QAPs and non-QAPs. Additionally, researchers probed for reasoning when participants gave limited responses. Questions were piloted to ensure clarity, appropriateness and sense-making. All interviews were conducted one-to-one apart from two, where two people from the same

organisation were present to provide comprehensive responses. Telephone/skype interviews were conducted by four members of the team (AG, PC, LM, MP), audio-recorded, and transcribed verbatim.

Data analysis

Data were analysed following the stages of framework analysis.[17] This approach was followed due to its retroductive inductive-deductive nature to test initial theory whilst identifying emerging Context, Mechanism and Outcomes (CMOs) configurations. For *familiarisation*, two researchers each read one transcript and then made notes to identify CMO configurations (Total: 5 researchers X 5 transcripts). All researchers met to discuss similarities and differences across the transcripts including recurrent CMO configurations. We then developed a coding framework, consistently applied to each transcript. Five researchers coded data using NVivo,[18] with frequent progress meetings.

Reflexivity

Prior to analysis, individual members completed a written reflexive exercise which highlighted prior dispositions towards the research, and were then discussed collectively. The team consisted of practitioners, academics and researchers from medical and social science background disciplines, with QA knowledge ranging novice-expert.[19] Team members were vastly experienced in qualitative methods[20 21] and had previously applied a realist lens to understand complex interventions in healthcare education.[22]

Patient and public involvement

Given the focus on regulator-medical provider interactions in this study, patients and the public were not involved in the design, data collection or data analysis

RESULTS

Participant details

Following ethical approval, we conducted interviews with 36 individuals representing 34 different organisations between July-September 2018 (table 1) to produce a considerable amount of data: 35 hours, 27 minutes. Interviews ranged between 48-88 minutes, mean = 63. The sample represented regional, national (England, Scotland, Northern Ireland, Wales) and international stakeholders within and outside medicine (Figure 3). There were 12 QAPs and 22 non-QAPs, with 27 (79%) of these from the UK and 7 (21%) international, representing Asia, North America and Europe. Participants often held senior roles such as chief executives and quality leads.

Table 1. Demographic information of the participant organisations

Key: QA=Quality Assurance partner, Health=health organisation, Med=medicine organisation, UG=undergraduate, PG=postgraduate, RC=royal college, INT=international based, UK=United Kingdom based

Interview number	ID	QA partner (Y/N)	Profession Sector	Location (coverage)
1	HealthMedINT1	N	Health, medicine	International
2	QAUG2	Y	Undergrad	UK
3	QARC3	Y	Royal College	UK
4	OtherprofessionUK4	N	Other Profession	UK
5	QAUG5	Y	Undergrad	UK
6	HealthMedUK6	N	Health, medicine	UK

7	HealthMedUK7	N	Health, medicine	UK
8	EducationUK8	N	Education	UK
9	HealthNon-MedUK9	N	Health, non-medicine	UK
10	QAPG10	Y	Postgrad	UK
11	HealthNon-MedINT11	N	Health, non-medicine	International
12	QAPG12	Y	Postgrad	UK
13	HealthMedINT13	N	Health, medicine	International
14	QAPG14	N	Health, medicine	UK
15	QAUG15	Y	Undergrad	UK
16	QAPG16	Y	Postgrad	UK
17	EducationUK17	N	Education	UK
18	HealthNon-MedUK18	N	Health, non-medicine	UK
19	QAUG19	Y	Undergrad	UK
20	HealthMedINT20	N	Health, medicine	International
21	HealthMedUK21	N	Health, medicine	UK
22	OtherprofessionUK22	N	Other profession	UK
23	HealthMedUK23	N	Health, medicine	UK
24	QARC24	Y	Royal College	UK
25	EducationUK25	N	Education	UK
26	HealthMedUK26	N	Health, medicine	UK
27	HealthNon-MedUK27	N	Health, non-medicine	UK
28	QAPG28	Y	Postgrad	UK
29	QAUG29	Y	Undergrad	UK
30	HealthMedINT30	N	Health, medicine	International
31	EducationUK31	N	Education	UK
32	EducationINT32	N	Education	International
33	QAPG33	Y	Postgrad	UK
34	HealthMedINT34	N	Health, medicine	International

Figure 3. Stakeholder groups

[Insert figure 3 here]

Main findings

We present findings which verify, refute and challenge our initial programme theory (themes 1-4), leading to the development of our modified programme theory. Contexts, mechanisms and outcomes are labelled as [C], [M], and [O] respectively. CMO configurations, resources and responses are identified and illustrated across themes. We found that depending on context, the same interventions triggered a range of mechanisms leading to positive or negative outcomes.

Theme 1) Quality standards

Importantly, standards defined the level at which a provider needs to function to reach certain outcomes, e.g. meeting minimum standards. Key mechanisms triggered by co-construction of setting standards included compliance, clarification, flexibility and adherence.

> *I think, a very positive element [O], is that it [standards] has allowed UK medical schools [C] within the framework to differ in how they implement that framework…So I think as well as having the rigour of what must be done, it allows for a degree of flexibility [M]. (HealthMedINT1)*

Undergraduate and postgraduate QA partners responded to the regulators standards by inclusion in their own policies: "We reference them [the standards] in our…internal policies" (QAUG19). The standards also provided QAP's with leverage to push forward changes at institutions "now [the standards] look at the environment rather than just specific areas like clinical supervision" (QAPG10) However, the standards had unintended consequences as their presence sometimes created confusion particularly in the postgraduate context where there was scepticism about the lack of clarity. In some instances, organisations had their own standards to assess educational quality, resulting in confusion "I suspect [postgraduate organisation] ignores them [standards] because they've come up with their own quality framework (QAPG33)".

> *Our biggest concerns [O] really are not so much the standards as the sort of processes that by which the GMC [C] will check that our curricula are... comply with those standards [M]. And I think on that, you know, looking back now I can see that there was a certain ambiguity [M] in how the GMC were going to approach this and I'm not sure that they ever resolved it as the standards were being developed [O]. (QARC3)*

Standards that are overly prescriptive, rigid and inflexible prevent providers from being adaptable to need and innovation. Conversely, less binding standards triggered mechanisms of ambiguity, openness and flexibility creating too much variation in education across contexts and producing new risks to quality. Misalignment between different quality standards caused frustration. For example, a LEP was deemed to be clinically outstanding but was also found to be inadequate for educational quality by a different regulator. Local pressures were seen to inhibit postgraduate partners' abilities to follow standards rigidly, suggesting that in the 'real world', applicability of standards was sometimes questionable:

> *A lot of LEPs take our students [C], but they [LEPs] can quite readily tell us [medical school] to take them away as well [O], if we're very strict with them about meeting certain standards and certain criteria [M]. (QAUG5)*

Theme 2) Sanctions and approvals

We identified that organisational culture affects approaches to sanctions, and so an 'acceptable sanction' was contingent on risk. However in different contexts, should supportive measures be inadequate then the most severe sanction of withdrawn approval should remain. The "ultimate sanction of power" (QAPG28) was regarded to fulfil its intended consequence i.e. to protect patient safety, but also had unintended consequences to reinforce the medical regulator's authority and subsequently motivate providers to address problems. There was a firm belief that a severe sanction should rarely need to be enforced if other QA components (e.g. self-assessment) are effective.

> *It's [sanctions] a bit of a lightning rod situation, but I think it [closing medical school] should remain as the ultimate sanction [O]…If trust management realised for example that they wouldn't lose their trainees as a result of not providing a safe and effective training environment [M]… I think [it would] slip further down their list of priorities [O]. (QARC24)*

The effectiveness of regulatory approvals in the undergraduate and postgraduate context varied. In undergraduate, it was described as time-consuming examining both curriculum and staff capabilities. The thoroughness "enabled them [regulator] to make a decision on our suitability to proceed" (QAUG29). However, in the postgraduate context the process and need for QA was not as clear. For instance, "what if

the trainee goes for one week, but it's only one week out of a one year placement, do they need to get that site approved?" (QAPG12). Non-QAPs felt that it was important that mechanisms were in place to periodically review approvals.

> *We don't link approvals and quality very strongly [C]…we go to the GMC and we say, can we put some doctors here please? And the GMC go, yes. But there's an implication in doing that that because we're asking, we're going to quality manage that particular set of placements [M]. And we do, but not explicitly and not formally [O]. (QAPG16)*

Theme 3) Collecting information: Visits, monitoring and self-assessment

Institutional visits was positioned as a key component as it triggered internal reviews and reflection, motivating organisations to improve quality. Working collaboratively engendered trust with open and honest dialogue, which was considered crucial in effecting change: "I was prepared to be completely open and honest with the GMC…If [visits] are going to be effective, relationship building is actually more important than what you're doing collecting evidence" (HealthMedUK21). Meaningful dialogue and collaboration were important and that was achieved through having high quality, "respected" (EducationINT32) trained visitor teams.

The QAF includes a range of monitoring data collection processes such as: data from the national training survey (NTS); monitoring including enhanced monitoring; self-assessment; and visits. The NTS surveys all doctors in training which facilitated increased transparency, accountability and risk-identification. The resources provided by the survey could lead to invaluable outcomes to examine training differences, make evidence-informed decisions and pinpoint training issues. However, multiple sources of data were sometimes regarded as conflictual, obscuring the overall picture of education quality.

> *At the moment, they're [GMC] looking [at] trainee burnout [C]. So they're generating all this data at the moment and I don't think they're clear about what they're going to do with it [M], and my concern is they will just dump it on us for us to fix, and I don't think we can [O]. (QAPG10)*

The component of requiring self-assessment triggered many different mechanisms in different contexts. For regulators, it generated reflective internally-led processes, "a really fundamental part of what we do, and we place a massive... emphasis on that" (EducationUK17). The reasons for this were around connectivity between regulators and providers, "this is the way [self-assessment] an institution connects itself with given standards" (EducationINT32); "to work constructively with the provider…being the start of a peer review process" (HealthNon-MedINT11). Whereas, for those who were being quality assured, the formality of written self-assessment inhibited open disclosure as it was laborious and seen more as an audit than self-assessment, "I think you're more likely to hear genuine issues, genuine things that need to be fixed, if you speak to people informally and off the record" (QAPG6). Validity and reliability was raised, "It [self-assessment] forms part of it[assuring quality] and it's a very strong part of it, but I wouldn't necessarily use it [self-assessment] in isolation" (HealthMedUK23). Perceptions of lack of feedback from the regulator also undermined self-assessment.

> *We are encouraging of institutions identifying challenges [C]. So if an institution is very open and honest [M], even into what might be quite a delicate area, saying this has been a challenge for us [C] and we're working away on it and we're doing the following things. Provided that their plan of action is a good one and that it's being conducted in a timely manner that would be reported on in a positive light [O]. (HealthNonMedUK9)*

Theme 4) Reporting; accountability, dissemination, good practice

A patient safety outcome response identified from external reporting was to build public confidence, "I think the transparency in publications are important because it involves or it makes things clear and open to all stakeholders" (QAPG28). Publicly available outputs fostered accountability to illustrate how providers are low-risk thus requiring less scrutiny. Insufficient reporting and feedback (in terms of timeliness, quantity and quality) fostered outcomes of devaluing time and effort, and subsequent disengagement.

> *I think what having it public does, is it creates some pressure and accountability [M] on both the accreditor and the accrediting body [C] to focus on the outcomes and to show progress against conditions [O]. (HealthMedINT13)*

Risk context was also important to determine the effectiveness of intervention components. Informal partnerships were highlighted as critical to assuring quality *"Working in partnership with regulators was instrumental, it has a significant effect on driving change in trusts"* (HealthedUK21). When feedback mechanisms triggered included collaboration and openness (rather than accountability) this fostered informal working partnerships leading to a raft of positive outcomes including awareness, sharing knowledge and quality improvement. Rapport over time helped provider's develop trust to report concerns. Here the positioning of regulator-provider context, moved from accountability quality-checker to collaborative problem-solver.

> *The institution needs to take that genuine look at it, and spend the time genuinely evaluating and genuinely creating action plans…an institution that is good at critical self-reflection will tend to address problems before, or potential problems, before they become actual problems [O]. (EducationUK17)*

Spreading good practice was contextually variable, "What works for one school may not work for other [O] …So, you don't want people to blindly be saying oh, let's do that, because that's going to be good practice here [M]….Because education in programs do vary [C]" (*HealthNon-MedUK18).* Subsequently motivation reaction was low to implement changes based on other organisations' examples of good practice. Reports were deemed most helpful when they included action plans accessible to lay audiences.

DISCUSSION

Summary of findings

We undertook a realist evaluation to explore medical education QA undertaken by the GMC. We found that intervention components support or undermine QA for different organisations, and at different times in undergraduate and postgraduate contexts. We tested our initial programme theory (themes 1-4) to develop a modified programme theory (presented below) which articulates the need for components to be tailored proportionately to contexts.

Across the three-tier model we identified that the undergraduate and postgraduate context were influential. The leverage brokered by the regulator in the undergraduate context was often associated with directive features enabling local changes, whereas in contrast for the postgraduate context this power was often lost and diffused across education layers. Predominantly in the postgraduate context, interventions led to unintended consequences (e.g. organisation disengagement) if an intervention promoted adherence at the cost of autonomy, subsequently triggering a lack of motivation (theme 1). An underlying mechanism identified to ensure an inclusive approach to QA was partnership working (themes 3-4). In the undergraduate context, provider-regulator engagement was sometimes not present, typically when there was a lack of informal relationships. In theme 3, visits were identified as a component that could better foster partnerships; so long as they were conducted with integrity, meaning and purpose.

Collectively across themes 1-4, the QAF was commended for being comprehensive and enabling a broad understanding of an organisation's performance. Internally-led processes with organisations identifying and addressing their own challenges and deficiencies, when done well, promoted a sense of autonomy and accountability (theme 3). The main unintended QAF consequences fell broadly into two outcomes, those related to the overlap across the three tiers (themes 1-2) and those related to the regulatory burden associated with data-driven approaches with a lack of transparency on why and how data was used (themes 3-4). A blurring of roles and boundaries of multiple organisations between patient safety, medical education and training was identified.

Modified programme theory

The findings informed a modified programme theory to explain the underlying processes for the intended and unintended consequences of how the GMC quality assure education in various contexts (Figure 4). Across the three-tier model (quality assurance, quality management and quality control), the theory demonstrates that QAF components are enacted differently with implications for the mechanisms triggered leading to positive (e.g. effecting change, contextual application of standards, partnerships) and/or negative outcomes (limited compliance, resistance, overlap). The influence and power of the regulator was continually picked up across the components (themes 1-4) which triggered mechanisms including transparent reporting to promote quality improvement, effective communication, trust, and partnership working facilitating interactions between regulators, partners and providers. Proportionate reactions in the face of disclosing and identifying patient safety risks at an early stage were more likely to occur within a positive trusting regulator-provider context underpinned by openness. Likewise, an organisation that self-assessed critically was reported to give regulators confidence in the institution.

Figure 4. Modified programme theory to explain the underlying process for the intended and unintended consequences of quality assurance components

[Insert figure 4 here]

Relevance of findings and implications

The findings reinforce the quality assurance literature highlighting *trust* in fostering effective working relationships to enhance feedback.[1 5] We extend this further, and identified that early communication of emerging risks supports quality assurance and enhancement approaches through informal networks. Visits aid communication and build relationships, yet if lost may distance the regulator and undermine opportunities for partnership working. Informal communication provided a safe environment for providers to discuss concerns with the regulator, opposed to formal monitoring acting as a barrier.

Expanding the literature, we demonstrate that context must be considered in order for quality assurance to protect patients.[23] Risk is context dependent, and was perceived to be tangibly different across undergraduate and postgraduate contexts. Undergraduate medical settings were perceived as low risk and imply opportunities for greater tailoring and focus. The overlaps between quality assurance, quality management and quality control were apparent especially within the postgraduate setting with duplication and confusion of responsibilities. These findings align with a recent systematic review identifying features of failing healthcare organisations including conflicting missions, fragmented accountability and lack of collaboration[1].

Collectively, this supports the need to clarify structural quality processes and how organisations are intended to function collaboratively. In the analysis, risk-based visiting positioned the regulator as quality assurer rather than quality enhancer. Equally, effective assurance is often associated with suppressing

innovation.[24] Moreover, the role of self-assessment[24] posed a number of challenges in relation to purpose and autonomy. While institutional self-assessments can positively influence reactions to drive quality improvement, there are issues with validity, reliability and internal quality review.[19 24-27]

The power of the regulator impacted on the effectiveness of intervention components in multiple ways. The regulatory-burden associated with monitoring activities was considerable and disengagement ensued. Lack of feedback from the regulator was an important aetiological mechanism precipitating the situation. Similarly, negative consequences of approvals including cost, low staff morale, threats to organisation reputation, and the suppression of innovation through adhering to standards has been identified.[28 29] Without regulators addressing varying risk contexts, the proportionality of QA is imbalanced, leading to negative outcomes with regulators unable to effectively assure quality. Therefore, collectively considering a hybrid model of cyclical plus risk-based visiting may help to build provider relationships and drive improvement whilst ensuring minimum standards. Collective assurance and relationships should be encouraged so that regulators and providers can tackle issues conjointly. Flexibility in utilising other datasets within any collaborative work is a necessity and a clear stance on organisational remit and particularly boundaries, is anticipated to be a key mechanism in effective joint QA.

Strengths, limitations and future directions

To our knowledge this is the first robust study on education QA within the healthcare context, synthesising data from stakeholders. The study fills a gap as QA remains expensive, yet efficacy is largely unexplored. The study was conducted by an experienced multidisciplinary research team applying an innovative realist approach, underpinned by a sound team-based analysis. A somewhat surprising omission from our findings is a lack of attention to the mechanism of leadership.[1 23] The sample focussed on processes rather than delivery perhaps contributing to such omission. Moving forward, there is a need to conduct an economic review and consult with stakeholders into what data could be shared (e.g. National Training Survey, Care Quality Commission data) to understand links to intervention components. The findings have influenced the GMC's approach to QA which impacts on all medical students and doctors in training.[30]

Conclusions

This study used a realist methodology to reveal the intended and unintended consequences of components used by the GMC to quality assure medical education, and elucidated the mechanisms by which both are brought about. While uniform approaches are often in place, interventions need to be contextually tailored. Ongoing partnership working can enhance open disclosure to drive up education quality. This research has provided a modified programme theory to explicate how education providers and regulators can work more effectively together to uphold quality, and ultimately protect public safety.

Acknowledgements

We would like to thank the participants for sharing their insights, the GMC for facilitating recruitment, and Dr Graham Easton and UCL Design team for helping to produce the informational video.

Contributorship

AG, PC, LM and MP conceived the study design. PC and LM designed study materials and gained ethical approval. All authors collected data, contributed to data analysis. PC wrote the first draft of the manuscript with input from AG, MP, LK and LM. All authors approve the final manuscript.

Data sharing

We do not have ethical permission to share raw data.

Funding and conflict of interest

This study was commissioned by the GMC, project ID: GMC 822. AG was the principal investigator.

Ethics

This study was registered with UCL's data protection office on 06/06/18 and approved by UCL ethics on 15/06/18, project ID: 6281/003. All participants volunteered to take part and provided written consent. Data were anonymised with respect to individuals and institutions.

References

1. Vaughn VM, Saint S, Krein SL, et al. Characteristics of healthcare organisations struggling to improve quality: results from a systematic review of qualitative studies. *BMJ Qual Saf 2019*, *28*(1), pp.74-84.

2. General Medical Council. Promoting excellence 2019. Accessed on 16.07.19 from: https://www.gmc-uk.org/education/standards-guidance-and-curricula/standards-and-outcomes/promoting-excellence

3. Vroeijenstijn, A.I. Quality assurance in medical education. *Academic medicine: journal of the Association of American Medical Colleges* 1995, *70*(7 Suppl), pp.S59-67.

4. General Medical Council (2019). How we assure quality. Accessed on 10/04/2019 from: https://www.gmc-uk.org/education/how-we-quality-assure

5. Jones K & Tymms P. Ofsted's role in promoting school improvement: the mechanisms of the school inspection system in England, *Oxford Review of Education 2014*, 40(3), 315-330, DOI: 10.1080/03054985.2014.911726

6. Westerheijden DF, Stensaker B, Rosa MJ (eds). *Quality assurance in higher education: Trends in regulation, translation and transformation* (Vol. 20) 2007. Springer Science & Business Media.

7. Haji F, Morin MP, Parker K. Rethinking programme evaluation in health professions education: beyond 'did it work?'. *Medical education 2013*, *47*(4), pp.342-351.

8. Øvretveit J, Gustafson D. Evaluation of quality improvement programmes. *BMJ Qual Saf* 2002;11:270-275.

9. Pawson R, Greenhalgh T, Harvey G, et al. Realist review – a new method of systematic review designed for complex policy interventions, *Journal of Health Services Research & Policy 2005* 10(1): 21–34

10. Pawson R. *The Science of Evaluation: A Realist Manifesto*, SAGE Publications Ltd, London, 2013.

11. Jones AC, Shipman SA, Ogrinc G. Key characteristics of successful quality improvement curricula in physician education: a realist review BMJ Qual Saf 2015;24:77-88

12. Wong G, Westhorp G, Manzano A, et al. RAMESES II reporting standards for realist evaluations. *BMC medicine 2016*, *14*(1), p.96.

13. Griffin A, Crampton P, Mehdizadeh L, et al. Understanding stakeholder perspectives on the GMC's quality assurance of medical education and training. Final Report, 2018. Accessed 25.07.19 from: https://www.gmc-uk.org/-/media/documents/understanding-stakeholder-perspectives-on-the-gmcs-quality-assurance-of-medical-education-and-traini.pdf?la=en&hash=14529235D9E6C99AF9B5EDA82F6848CB34D33396

14. Jagosh J, Bush PL, Salsberg J et al. A realist evaluation of community-based participatory research: partnership synergy, trust building and related ripple effects. *BMC public health 2015*, *15*(1), p.725.

15. Manzano A. The craft of interviewing in realist evaluation. Evaluation. 2016 Jul;22(3):342-60.

16. Keren G, De Bruin WB. On the assessment of decision quality: Considerations regarding utility, conflict and accountability. Thinking: Psychological perspectives on reasoning, judgment and decision making. 2003 Sep 16;347.

17. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. *The qualitative researcher's companion* 2002; 573, 305-329.

18. NVivo qualitative data analysis software; QSR International Pty Ltd. Version 11, 2015.

19. Griffin A, McKeown A, Viney R, et al. Revalidation and quality assurance: the application of the MUSIQ framework in independent verification visits to healthcare organisations. *BMJ Open* 2017;7:e014121. doi:10.1136/bmjopen-2016-014121

20. Griffin A, Knight L, McKeown A, et al. A postgraduate curriculum for integrated care: a qualitative exploration of trainee paediatricians and general practitioners' experiences. *BMC Med Educ* 2019; *19*(1), p.8.

21. Rees CE, Crampton PES, Kent F, et al. Understanding students' and clinicians' experiences of informal interprofessional workplace learning: an Australian qualitative study. BMJ open 2018.1;8(4):e021238.

22. Ajjawi R, Crampton PES, Rees CE. What really matters for successful research environments? A realist synthesis. *Medical education 2018*, *52*(9), pp.936-950.

23. Kaplan HC, Provost LP, Froehle CM, & Margolis PA. The Model for Understanding Success in Quality (MUSIQ): building a theory of context in healthcare quality improvement. *BMJ Qual Saf 2012*, *21*(1), pp.13-20.

24. Lang DW. Self-regulation with rules: Lessons learned from a new quality assurance process for Ontario, *Quality Assurance in Education 2015*, 23:3, 216-232, https://doi.org/10.1108/QAE-09-2014-0046

25. Tackett S, Grant J & Mmari K. Designing an evaluation framework for WFME basic standards for medical education, *Medical Teacher 2016*, 38(3), 291-296, DOI:10.3109/0142159X.2015.1031737

26. Jones KL, Tymms P, Kemethofer D, et al. The unintended consequences of school inspection: the prevalence of inspection side-effects in Austria, the Czech Republic, England, Ireland, the Netherlands, Sweden, and Switzerland, *Oxford Review of Education 2017*, 43(6), 805-822, DOI: 10.1080/03054985.2017.1352499

27. Forrest C. Inspection and improvement in three further education colleges, *Research in Post-Compulsory Education 2015*, 20(3), 296-314, DOI: 10.1080/13596748.2015.1063273

28. de Paor C. The contribution of professional accreditation to quality assurance in higher education, *Quality in Higher Education 2016*, 22(3), 228-241, DOI: 10.1080/13538322.2016.1263925

29. Blouin D, Tekian A, Kamin C, et al. The impact of accreditation on medical schools' processes, *Medical Education* 2018, 52(2), 182-191

30. General Medical Council. Council Papers February 2019. Accessed on 17.07.19 from: https://www.gmc-uk.org/-/media/documents/council---agenda-and-papers-for-meeting-on-27-february-2019_pdf-78924314.pdf

Figure 1: The intervention: quality assurance framework (QAF)

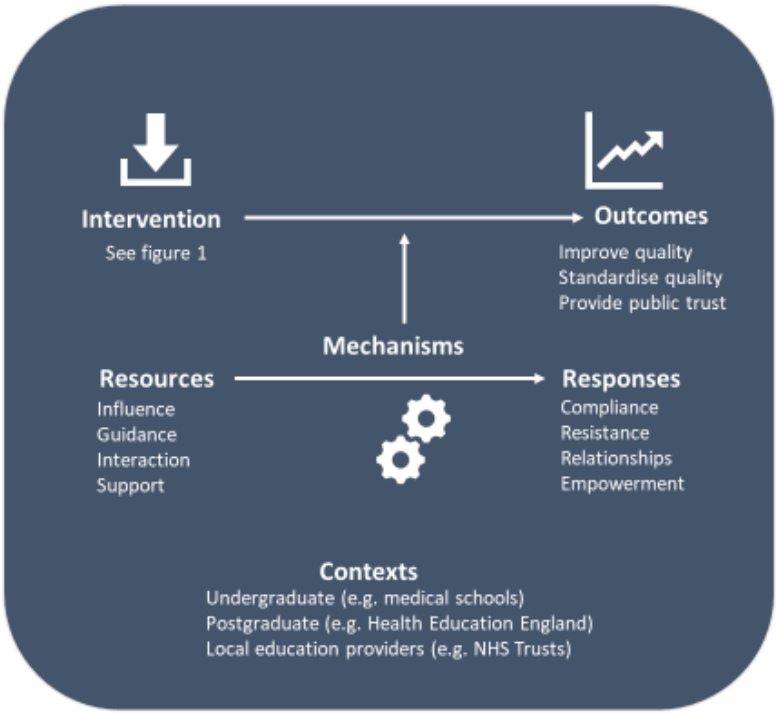


Figure 2. Initial programme theory

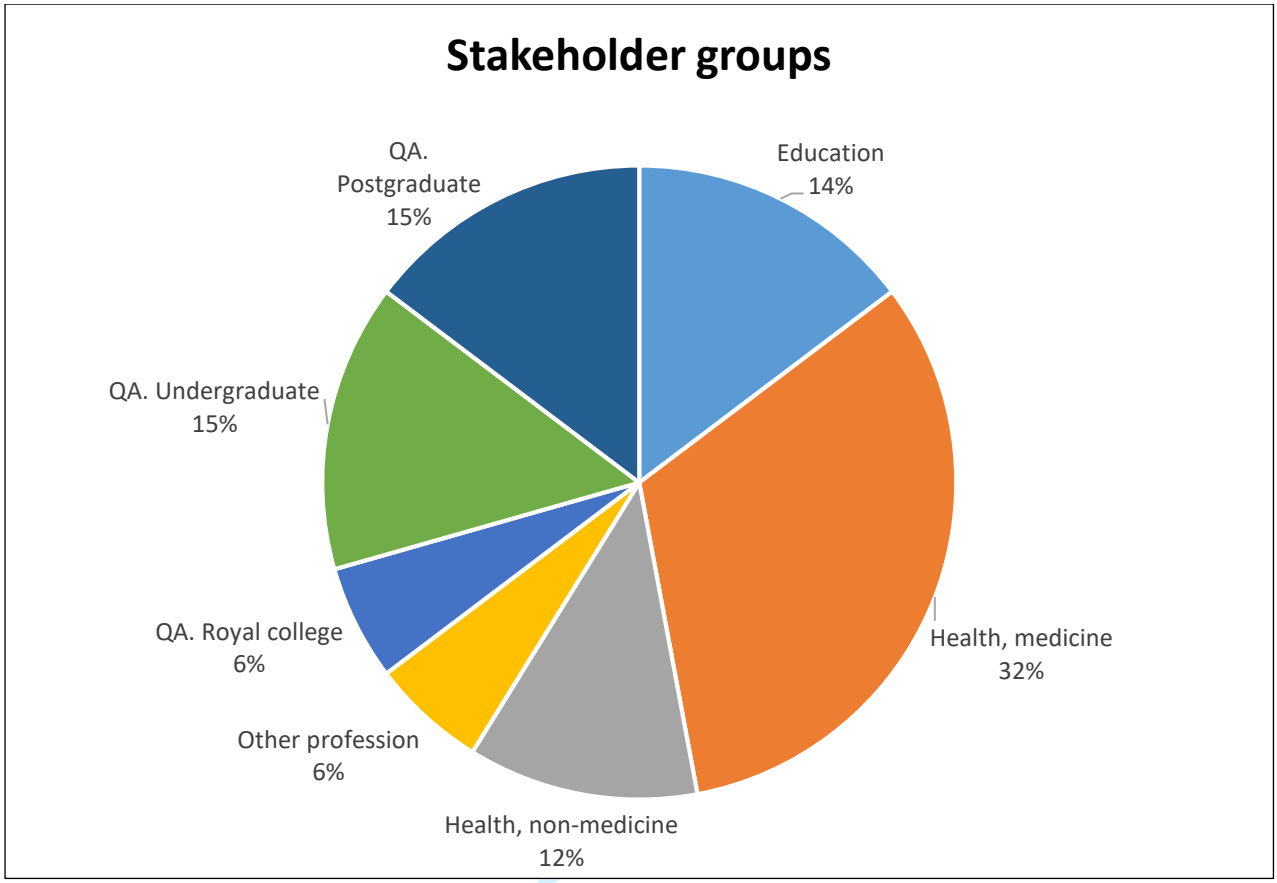


Figure 3. Stakeholder groups

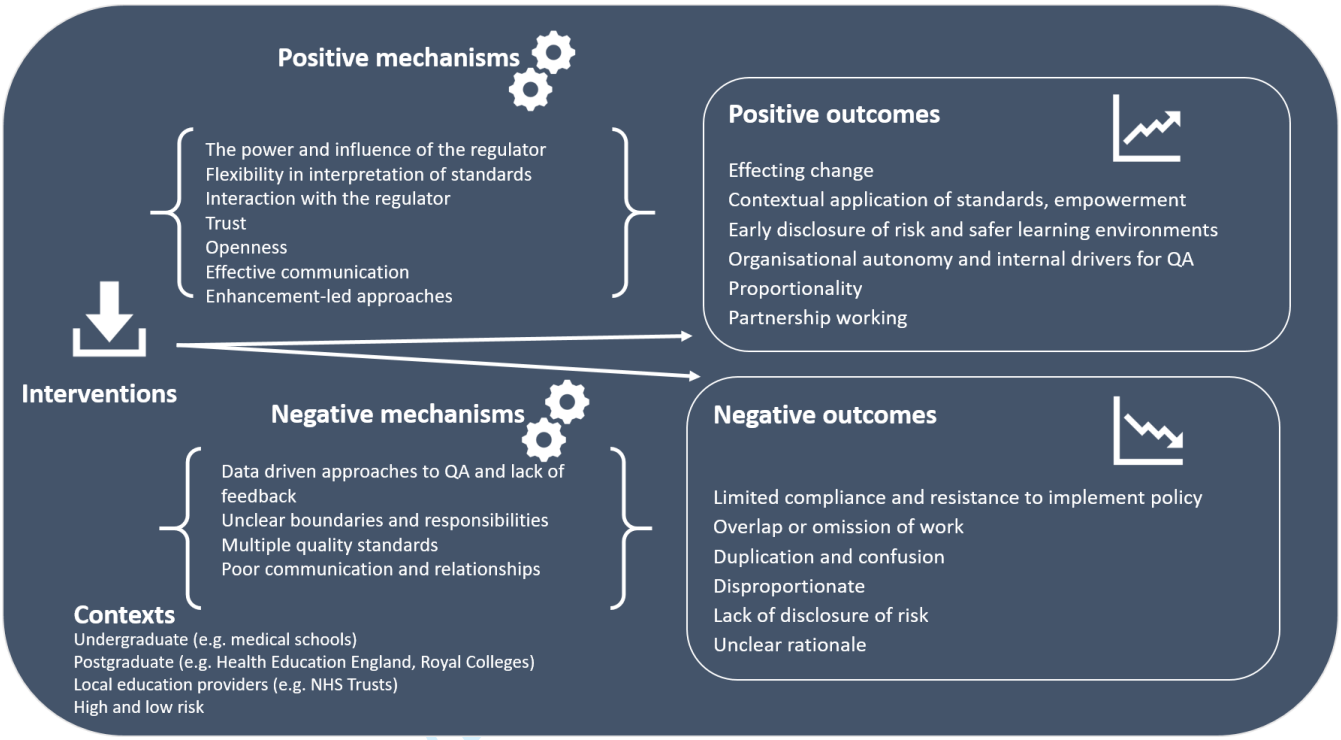


Figure 4. Modified programme theory to explain the underlying process for the intended and unintended consequences of quality assurance components

1. Quality Assurance partners interview schedule

Structure

Introduce and duration about 45 mins

Process

This interview will be recorded and analysed, looking for common themes that arise across the interviews.

Consent

A reminder that we will not personally name anyone who takes part in the study. Do you have any questions before we start the tape? Thank you for signing the consent form [or take verbal consent if required].

Introductory/ Background questions

- How long have you worked in your quality assurance role?
- Who else in your organisation works in a QA role?

General: GMC quality assurance framework overall

- Are there any aspects of GMC's quality assurance framework that you think are particularly effective, i.e. give you reassurance in their processes?
- Are there any aspects of the framework that you think are less effective or problematic in some way, i.e. do not assure you?

Focused: Specific aspects of GMC's quality assurance framework

I would now like to ask some questions about different components of the QA framework and be keen for you to share your experiences where relevant.

Standards

- Are the standards the right ones? Prompt any missing?
- Are the standards helpful or unhelpful in anyway?
- Has using the standards had any impact on your organisation?

Approvals

- What would be the advantages of making the GMC's approvals time limited?
- What would be the disadvantages of making the date GMC's approvals time limited?
- Do you think the GMC's approvals process is effective?

Monitoring:

- Is the GMC monitoring the right evidence to assess your organisations performance?

- What sources of evidence do you think give the GMC the best insight into your organisation?
- What other areas could they/should they monitor?
- Does monitoring have any impact on your organisation? Prompt: Positive/negative
- Turning to enhanced monitoring, some people would say that the GMC are overstepping their remit when they require postgraduate organisations to report training programs and local education providers to them, what are your thoughts?

Sharing evidence:

- How effective is the GMC at sharing evidence with you?
- Is there evidence that could be shared more effectively and how would that benefit your organisation?
- Is there any evidence that you feel should not be shared, in particular with other healthcare regulators?

Self-assessment:

The GMC requires annual self-assessment from the medical Royal colleges and medical schools but not the postgraduate organisations.

- Do you think self-assessment is a helpful process?
- Some hold the view that organisational self-assessment is not a reliable process, what do you think?
- Has the process of self-assessment resulted in any organisational change?

Visits:

- What purpose do you think visits to organisations have?
    - Prompt: What makes a visit effective?
    - Prompt: What are the important areas that visits should include?
- Most regulators are moving away from cyclical or scheduled visiting, towards entirely risk-based systems, however many GMC stakeholders believe the cyclical visits have many benefits and should be retained. What do you think?
- What would happen if the GMC did no visiting?

Reporting:

The GMC currently publishes long-form visit reports on its website, as well as information about enhanced monitoring cases, and data tools such as the NTS reporting tool and the progression reports.

- What do you think of the current QA reporting?
- Are there any negative consequences of reporting data on the website?
- Does your organisation use the reports in anyway?

Good practice:

The GMC aim to identify good practice across medical schools and postgraduate bodies and then publish this on their website.

- Is this useful to your organisation?
    - Prompt: positive aspects v negative

- Have you adopted any areas of good practice yourself?
- Some people would say more resources should go into quality enhancement rather than accountability. What are your views?

Fairness

- How can the GMC quality assure fairness in medical education and training?

Sanctions:

Sanctions mean withdrawing trainee doctors from the NHS or closing down medical schools which has a critical impact on healthcare.

- In the case of an underperforming training organisation that is currently failing to meet required standards what might be a proportionate sanction from the GMC that is not as extreme as withdrawing approval?
    - Prompts: The GMC visiting, publicly available rating scales, time bound approvals

Collective assurance

The GMC has committed to working more closely with other regulators to find efficiencies and reduce the regulatory burden on the service.

- What would be the advantages for your organisation in this approach?
- Would there be any disadvantages?
- Would sharing data enable the GMC to identify risk better?
- How practical would it be for your organisation to undertake joint visiting?
- Do you think the GMC's approach to QA is proportionate to the risks involved in medical education and training?
- Do you have any suggestions for improving the GMC's quality assurance processes?

Thank you for your time. Is there anything you would like to add that we haven't discussed?

Thank you

2. Quality Assurance non-partners interview schedule

Structure

Introduce and duration about 45 mins

Process

This interview will be recorded and analysed, looking for common themes that arise across the interviews.

Consent

A reminder that we will not personally name anyone who takes part in the study. Do you have any questions before we start the tape? Thank you for signing the consent form [or take verbal consent if required].

Introductory/ Background questions

- Can you briefly explain the context in which your organisation is involved in QA
- What is your specific role?
- How long have you worked in your quality assurance role?
- Who else in your organisation works in a QA role?

General: GMC quality assurance framework overall

- Are there any aspects of GMC's quality assurance framework that you think are particularly effective, i.e. give you reassurance in their processes?
- Are there any aspects of the framework that you think are less effective or problematic in some way, i.e. do not assure you?

Focused: Specific aspects of GMC's quality assurance framework

I would now like to ask some questions about different components of the QA framework and be keen for you to share your experiences where relevant.

Standards

- Are the standards the right ones?
    - Prompt: Any missing?
- Are the standards helpful or unhelpful in anyway?

Approvals

- Do you think the GMC's approvals process is effective?
- What would be the advantages of making the GMC's approvals time limited?
- What would be the disadvantages of making the date GMC's approvals time limited?

Monitoring

- Is the GMC monitoring the right evidence to assess organisational performance?
    - Prompt: What other areas could they/should they monitor?
- Do you think monitoring has any impact on organisation performance?
    - Prompt: Positive/negative
- Turning to enhanced monitoring, some people would say that the GMC are overstepping their remit when they require postgraduate organisations to report training programs and local education providers to them, what are your thoughts?
- How does your organisation use monitoring?
- Do you have a model for triangulating predicting risk?

Sharing evidence

- How could the GMC improve sharing its evidence?
    - Prompt: Between regulator to regulator; between regulators to QA partners?
- Is there other evidence that could be shared?
    - Prompt: Is there any evidence that you feel should not be shared?

Self-assessment

The GMC requires annual self-assessment from the medical Royal colleges and medical schools but not the postgraduate organisations.

- Do you think self-assessment is a helpful process?
- Some hold the view that self-assessment is not a reliable process, what do you think?
- What is your organisations approach to self-assessment?
- Has the process of self-assessment resulted in any organisational change?

Visits

- What purpose do you think visits to organisations have?
    - Prompt: What makes a visit effective?
    - Prompt: What impact do you think organisational visits have?
- Most regulators are moving away from cyclical or scheduled visiting, towards entirely risk-based systems, however many GMC stakeholders believe the cyclical visits have many benefits and should be retained. What do you think?
- How can visits give greater assurance of quality?
- What would happen if the GMC did no visiting?

Reporting

The GMC currently publishes long-form visit reports on its website, as well as information about enhanced monitoring cases, and data tools such as the NTS reporting tool and the progression reports.

- What do you think of the GMC approach to reporting?
- How does your organisation report on performance?
    - Prompts: strengths and weaknesses?

Good practice

The GMC aim to identify good practice across medical schools and postgraduate bodies and then publish this on their website.

- What do you think of the GMCs approach to sharing best practice?
- Some people would say more resources should go into quality enhancement rather than accountability. What are your views?
- What is your organisations approach to this?

Fairness

- How can the GMC quality assure fairness in medical education and training?

Sanctions

In the GMC's context, sanctions mean withdrawing trainee doctors from the NHS or closing down medical schools which has a critical impact on healthcare.

- Does your organisation have any advice or experience of imposing meaningful sanctions that would not be considered as extreme as the GMC's approach?

Collective assurance

The GMC has committed to working more closely with other regulators to find efficiencies and reduce the regulatory burden on the service.

- Is your organisation involved in joint visits? If so, what would be the advantages/ disadvantages for your organisation in this approach?
- Do you sharing data with other organisations?
- How practical would it be for your organisation to undertake joint visiting?
- Do you think the GMC's approach to QA is proportionate to the risks involved in medical education and training?
- Do you have any suggestions for improving the GMC's quality assurance processes?

Thank you for your time. Is there anything you would like to add that we haven't discussed?

Thank you

RAMESES 2 reporting checklist

Source: Wong, G., Westhorp, G., Manzano, A., Greenhalgh, J., Jagosh, J. and Greenhalgh, T., 2016. RAMESES II reporting standards for realist evaluations. *BMC medicine*, *14*(1), p.96.

TITLE			**Reported in document**	**Section, Page(s) in document**
1		In the title, identify the document as a realist evaluation	Y	Title, pg 1
SUMMARY OR ABSTRACT				
2		The abstract or summary should include brief details on: the policy, programme or initiative under evaluation; programme setting; purpose of the evaluation; evaluation question(s) and/or objective(s); evaluation strategy; data collection, documentation and analysis methods; key findings and conclusions Where journals require it and the nature of the study is appropriate, brief details of respondents to the evaluation and recruitment and sampling processes may also be included Sufficient detail should be provided to identify that a realist approach was used and that realist programme theory was developed and/or refined	Y	Abstract, pg 2
INTRODUCTION				
3	Rationale for evaluation	Explain the purpose of the evaluation and the implications	Y	Introduction pg. 3-4

TITLE			Reported in document	Section, Page(s) in document
		for its focus and design		
4	Programme theory	Describe the initial programme theory (or theories) that underpin the programme, policy or initiative	Y	Methodology pg 4
5	Evaluation questions, objectives and focus	State the evaluation question(s) and specify the objectives for the evaluation. Describe whether and how the programme theory was used to define the scope and focus of the evaluation	Y	Introduction, study aim, pg 3-5
6	Ethical approval	State whether the realist evaluation required and has gained ethical approval from the relevant authorities, providing details as appropriate.	Y	Ethical approval, pg 5 All participants volunteered to take part and provided written consent.
METHODS				
7	Rationale for using realist evaluation	Explain why a realist evaluation approach was chosen and (if relevant) adapted	Y	Methods, conceptual framework, pg 4-5
8	Environment surrounding the evaluation	Describe the environment in which the evaluation took place	Y	Methods, conceptual framework, data collection, pg 4-6
9	Describe the programme policy, initiative or product evaluated	Provide relevant details on the programme, policy or initiative evaluated	Y	Introduction, pg 3-4; Methods, conceptual framework, Pg 5-6
10	Describe and	A description and justification	Y	Introduction, pg

TITLE			Reported in document	Section, Page(s) in document
	justify the evaluation design	of the evaluation design (i.e. the account of what was planned, done and why) should be included, at least in summary form or as an appendix, in the document which presents the main findings. If this is not done, the omission should be justified and a reference or link to the evaluation design given. It may also be useful to publish or make freely available (e.g. online on a website) any original evaluation design document or protocol, where they exist		4; Methods, conceptual framework, Pg 5-6
11	Data collection methods	Describe and justify the data collection methods – which ones were used, why and how they fed into developing, supporting, refuting or refining programme theory Provide details of the steps taken to enhance the trustworthiness of data collection and documentation	Y	Methods, data collection, pg 4-5
12	Recruitment process and sampling strategy	Describe how respondents to the evaluation were recruited or engaged and how the sample contributed to the development, support, refutation or refinement of programme theory	Y	Methods, sampling, recruitment, pg 4-5
13	Data analysis	Describe in detail how data	Y	Methods,

TITLE			Reported in document	Section, Page(s) in document
		were analysed. This section should include information on the constructs that were identified, the process of analysis, how the programme theory was further developed, supported, refuted and refined, and (where relevant) how analysis changed as the evaluation unfolded		analysis, reflexivity, pg 5
RESULTS				
14	Details of participants	Report (if applicable) who took part in the evaluation, the details of the data they provided and how the data was used to develop, support, refute or refine programme theory	Y	Results, participant details, pg 5
15	Main findings	Present the key findings, linking them to contexts, mechanisms and outcome configurations. Show how they were used to further develop, test or refine the programme theory	Y	Results, main findings, pg 5-9
DISCUSSION				
16	Summary of findings	Summarise the main findings with attention to the evaluation questions, purpose of the evaluation, programme theory and intended audience	Y	Discussion, summary of key findings, pg 9-11
17	Strengths, limitations and future directions	Discuss both the strengths of the evaluation and its limitations. These should include (but need not be limited	Y	Discussion, Strengths, limitations and future directions, pg

TITLE			Reported in document	Section, Page(s) in document
		to): (1) consideration of all the steps in the evaluation processes; and (2) comment on the adequacy, trustworthiness and value of the explanatory insights which emerged The particular implications arising from the realist nature of the findings should be reflected in these discussions		11-12
18	Comparison with existing literature	Where appropriate, compare and contrast the evaluation's findings with the existing literature on similar programmes, policies or initiatives	Y	Discussion, Comparison with existing literature and implications, pg 10-12
19	Conclusion and recommendations	List the main conclusions that are justified by the analyses of the data. If appropriate, offer recommendations consistent with a realist approach	Y	Discussion, implications, conclusions, pg 10-12
20	Funding and conflict of interest	State the funding source (if any) for the evaluation, the role played by the funder (if any) and any conflicts of interests of the evaluators	Y	Funding and competing interests, pg 12 This study was externally commissioned by the GMC

BMJ Open

A realist evaluation of UK medical education quality assurance

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-033614.R1
Article Type:	Original research
Date Submitted by the Author:	16-Oct-2019
Complete List of Authors:	Crampton, Paul; Hull York Medical School, Health Professions Education Unit; UCL, Research Department for Medical Education Mehdizadeh, Leila; UCL, Research Department for Medical Education Page, Michael; QMUL, Institute for Health Sciences Education Knight, Laura; UCL, Research Department for Medical Education Griffin, Ann; University College London, Medical School
Primary Subject Heading:	Medical education and training
Secondary Subject Heading:	Medical education and training, Qualitative research
Keywords:	QUALITATIVE RESEARCH, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, MEDICAL EDUCATION & TRAINING

Title page

Title: A realist evaluation of UK medical education quality assurance

Authors: Paul E.S. Crampton[1,3] , Leila Mehdizadeh[1] , Michael Page[4] , Laura Knight[1] , Ann Griffin[2]

Address affiliations:

[1] Research Department of Medical Education, UCL Medical School, Royal Free Hospital, Room GF/664, London, NW3 2PF, UK

[2] Research Department of Medical Education, UCL Medical School, The Directorate, 74 Huntley Street, London, WC1E 6AU, UK

[3] Hull York Medical School, York University, John Hughlings Jackson Building, University Rd, Heslington, York YO10 5DD, UK

[4] Institute for Health Sciences Education, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, Room 3.15, 3rd Floor, Garrod Building, Turner Street, Whitechapel, London, E1 2AD UK

Corresponding author: Ann Griffin, email: a.griffin@ucl.ac.uk, Research Department of Medical Education, UCL Medical School, The Directorate, 74 Huntley Street, London, WC1E 6AU, UK

Word count: 4052

Key words (MESH): Quality assurance, health care; Qualitative research; Medical Education

ABSTRACT

Objectives: The aim of the study was to explore what components of the General Medical Council's (GMC) quality assurance framework work, for whom, in what circumstances and how?

Setting: UK undergraduate and postgraduate medical education and training.

Participants: We conducted interviews with a stratified sample of 36 individuals. This included those who had direct experiences, as well as those with external insights, representing local, national and international organisations within and outside medicine.

Intervention: The GMC quality assure education to protect patient and public safety utilising complex intervention components including meeting standards, institutional visits and monitoring performance. However, the context in which these are implemented matters. We undertook an innovative realist evaluation to test an initial programme theory. Data were analysed using framework analysis.

Results: Across components of the intervention, we identified key mechanisms including: transparent reporting to promote quality improvement; dialogic feedback; partnership working facilitating interactions between regulators and providers, and; role clarity in conducting proportionate interventions appropriate to risk. The GMC's framework was commended for being comprehensive and enabling a broad understanding of an organisation's performance. Unintended consequences included confusion over roles and boundaries in different contexts which often undermined effectiveness.

Conclusions: This realist evaluation substantiates the literature and reveals deeper understandings about quality assuring medical education. While standardised approaches are implemented, interventions need to be contextually proportionate. Routine communication is beneficial to verify data, share concerns and check-risk; however, ongoing partnership working can foster assurance. The study provides a modified programme theory to explicate how education providers and regulators can work more effectively together to uphold education quality, and ultimately protect public safety. The findings have influenced the GMC's approach to quality assurance which impacts on all medical students and doctors in training.

STRENGTHS AND LIMITATIONS

• Quality assurance of medical education remains an expensive process yet there is very limited research to understand its effectiveness

• This study, underpinned by a sound team-based reflexive analysis, is the first in-depth realist evaluation of quality assurance in the healthcare context which is responsible for all medical doctors and students in training

• We found that depending on context, the same interventions triggered a range of mechanisms leading to positive or negative outcomes

• The study is specific in its UK focus however we collected qualitative data from a large number of UK based and international expert stakeholders within and outside medicine, enhancing the transferability of our findings

• While standardised approaches are implemented, interventions need to be contextually proportionate. Routine communication is beneficial to verify data, share concerns and check-risk; however, ongoing partnership working can better foster assurance to protect patient safety

INTRODUCTION

Rationale

Healthcare regulators quality assure education and protect public safety utilising complex intervention components including setting standards, institutional visits and monitoring performance. However, the context in which these components are implemented matters.[1-3] Within undergraduate and postgraduate medical education and training, the taught curriculum integrates with workplace-based experiential learning. Consequently education environments range from structured classrooms in university contexts to clinical placements within shifting healthcare structures across primary and secondary care. Therefore, the challenge for regulators is to mediate the quality of education and training across these spaces in order to assure the public that education and training is safe, that all medical students are prepared for practice and that all trainee doctors are fit to practise.

In the UK context, the General Medical Council (GMC) regulator work closely with other organisations to secure its standards, using a three-tier model. The GMC (Tier 1, quality assurance), has an overarching responsibility to hold undergraduate and postgraduate training bodies to account for meeting standards. These bodies (Tier 2, quality management) organise, manage, commission, and sometimes deliver medical education. They also manage quality in local education providers (LEPs), where students and trainees are placed, such as trusts, health boards, general practices, and other clinical settings. The LEPs (Tier 3, quality control), have processes to ensure satisfactory clinical placements, and that their organisation provides an appropriate learning environment. Medical royal colleges work with the GMC to ensure their curricula and assessments are fit for purpose, inform specialty and postgraduate programme delivery, and have local systems to support training.

The GMC has a multifaceted intervention to examine the quality of medical education and training provision, known as the quality assurance framework (QAF, figure 1). The intervention includes the following components: setting standards, approving education settings, monitoring activities including self-assessment and enhanced monitoring, visits, sharing evidence with other regulators and identifying good practice.[4] The QAF operates across the three-tier model i.e. between the regulator (e.g. GMC), organisational provider bodies (e.g. medical schools) and local service delivery organisations (e.g. hospitals, general practice).

Figure 1: The intervention: quality assurance framework (QAF)

[Insert figure 1 here]

A range of approaches can be implemented to assure education quality, from heavily arbitrated measures to informal uncontrolled processes. Existing education Quality Assurance (QA) research is sparse and tends to come from the field of school-based and higher education.[5 6] Whilst exploring the mechanisms of action of school inspections,[5] a theory stated that regulatory activities associated with improvement include: setting standards; the provision of feedback; employing a system of sanctions and rewards; monitoring schools by the collection of information and public accountability. However, more research is needed to understand within the healthcare context how quality assurance can protect patients. Despite large amounts of resources dedicated to education QA, there remains a lack of clear evidence. QA takes place within varied and complex social environments. For this reason, the same intervention can impact on individuals, teams and organisations in different ways.[7 8] Although there are intended consequences explicit in the QAF design, the implicit underlying drivers of these are not clear.

Specific aim

The study aim was to explore what components of the GMC's quality assurance framework work, for whom, in what circumstances and how?

METHODOLOGY

Conceptual framework: realist evaluation

We chose realist methodology because of a focus on four theoretically constructed and inter-related questions: what works, for whom, in what circumstances, and how?[9-11] This results in generative causation, about how QAF components operate, offering an assessment of *whether* they work, as well as *why*. In the results, we explore the complex configuration links between contexts (where, when and with whom the activity takes place), components (different activities applied to assure quality), mechanisms (underlying processes for why the activity is/is not effective), and outcomes (intended and unintended consequences).[12] Our methods follow the RAMESES 2 (realist and meta-narrative evidence synthesis: evolving standards)[12] reporting guidelines (for full report see Griffin).[13]

Initial programme theory

We developed an initial programme theory (figure 2) based on existing literature, the GMC's approach to QA and research team insight (see reflexivity). We positioned the QAF as consisting of various components which we then explored to answer our study aim. Each component triggers multiple responses when applied in certain contexts with underpinning resources. Our programme theory postulated that within undergraduate, postgraduate and local education provider contexts, the QAF led to improved quality and protected the public through exploiting regulatory influence, guidance and supporting organisations, leading to compliance, resistance, relationships and empowerment.

Figure 2. Initial programme theory

[Insert figure 2 here]

Sampling and recruitment

Stratified purposive sampling was used to test our theory with stakeholders. We targeted those who were familiar with the QAF, labelled in the study as quality assurance partners (QAPs); as well as those with outsider perceptions of how the framework is positioned in society (e.g. other regulators) and broader regulation contexts (e.g. education). These spanned organisations both inside and outside healthcare, including internationally; collectively labelled as non-quality assurance partners (non-QAPs). See table 1 for the list of QAPs.

Our realist position acknowledged that stakeholders each had partial knowledge of the intervention, therefore to fully explore research questions we included policy makers, implementers and recipients.[14] Following email invitations, non-responders received two reminders. Aligned to realist evaluation,[15] participants were given the opportunity to review project materials prior to participation via a 15-minute informational video.

Data collection

To test our programme theory we undertook semi-structured interviews to explore underlying processes triggered by QA components (Supplementary file 1). We tested a number of candidate theories to explore the underlying ways in which the intervention was intended to be successful. For example, to test theories on the impact of generalisability and utility[16] we asked: "Some hold the view that organisational self-assessment is not a reliable process, what do you think?", and "What would happen if the medical regulator (GMC) did no organisational visiting?" We designed questions with different foci appropriate to participants,[15] for QAPs and non-QAPs. Additionally, researchers probed for reasoning when participants

gave limited responses. Interview questions were piloted with a QA manager within our own medical school. Very minor changes were required to enhance clarity, appropriateness and sense-making. All interviews were conducted one-to-one apart from two, where two people from the same organisation were present to provide comprehensive responses. Telephone/skype interviews were conducted by four members of the team (AG, PC, LM, MP), audio-recorded, and transcribed verbatim. Participants were geographically dispersed across the world therefore face-to-face interviews were not feasible.

Data analysis

Data were analysed following the stages of framework analysis.[17] This approach was followed due to its retroductive inductive-deductive nature to test initial theory whilst identifying emerging Context, Mechanism and Outcomes (CMOs) configurations. For *familiarisation*, two researchers each read one transcript and then made notes to identify CMO configurations (Total: 5 researchers X 5 transcripts). All researchers met to discuss similarities and differences across the transcripts including recurrent CMO configurations. We then developed a coding framework, consistently applied to each transcript. Five researchers coded data using NVivo,[18] with frequent progress meetings.

Reflexivity

Prior to analysis, individual members completed a written reflexive exercise which highlighted prior dispositions towards the research, and were then discussed collectively. The team consisted of practitioners, academics and researchers from medical and social science background disciplines, with QA knowledge ranging novice-expert.[19] Team members were vastly experienced in qualitative methods[20 21] and had previously applied a realist lens to understand complex interventions in healthcare education.[22]

Patient and public involvement

Given the focus on regulator-medical provider interactions in this study, patients and the public were not involved in the design, data collection or data analysis

RESULTS

Participant details

Following ethical approval, we conducted interviews with 36 individuals representing 34 different organisations between July-September 2018 (table 1) to produce a considerable amount of data: 35 hours, 27 minutes. Interviews ranged between 48-88 minutes, mean = 63. The sample represented regional, national (England, Scotland, Northern Ireland, Wales) and international stakeholders within and outside medicine (Figure 3). There were 12 QAPs and 22 non-QAPs, with 27 (79%) of these from the UK and 7 (21%) international, representing Asia, North America and Europe. Participants often held senior roles such as chief executives and quality leads.

Table 1. Demographic information of the participant organisations

Key: QA=Quality Assurance partner, Health=health organisation, Med=medicine organisation, UG=undergraduate, PG=postgraduate, RC=royal college, INT=international based, UK=United Kingdom based

Interview number	ID	QA partner (Y/N)	Profession Sector	Location (coverage)
1	HealthMedINT1	N	Health, medicine	International
2	QAUG2	Y	Undergrad	UK

3	QARC3	Y	Royal College	UK
4	OtherprofessionUK4	N	Other Profession	UK
5	QAUG5	Y	Undergrad	UK
6	HealthMedUK6	N	Health, medicine	UK
7	HealthMedUK7	N	Health, medicine	UK
8	EducationUK8	N	Education	UK
9	HealthNon-MedUK9	N	Health, non-medicine	UK
10	QAPG10	Y	Postgrad	UK
11	HealthNon-MedINT11	N	Health, non-medicine	International
12	QAPG12	Y	Postgrad	UK
13	HealthMedINT13	N	Health, medicine	International
14	QAPG14	N	Health, medicine	UK
15	QAUG15	Y	Undergrad	UK
16	QAPG16	Y	Postgrad	UK
17	EducationUK17	N	Education	UK
18	HealthNon-MedUK18	N	Health, non-medicine	UK
19	QAUG19	Y	Undergrad	UK
20	HealthMedINT20	N	Health, medicine	International
21	HealthMedUK21	N	Health, medicine	UK
22	OtherprofessionUK22	N	Other profession	UK
23	HealthMedUK23	N	Health, medicine	UK
24	QARC24	Y	Royal College	UK
25	EducationUK25	N	Education	UK
26	HealthMedUK26	N	Health, medicine	UK
27	HealthNon-MedUK27	N	Health, non-medicine	UK
28	QAPG28	Y	Postgrad	UK
29	QAUG29	Y	Undergrad	UK
30	HealthMedINT30	N	Health, medicine	International
31	EducationUK31	N	Education	UK
32	EducationINT32	N	Education	International
33	QAPG33	Y	Postgrad	UK
34	HealthMedINT34	N	Health, medicine	International

Figure 3. Stakeholder groups

[Insert figure 3 here]

Main findings

We present findings which verify, refute and challenge our initial programme theory (themes 1-4), leading to the development of our modified programme theory. Contexts, mechanisms and outcomes are labelled as [C], [M], and [O] respectively. CMO configurations, resources and responses are identified and illustrated across themes. We found that depending on context, the same interventions triggered a range of mechanisms leading to positive or negative outcomes.

Theme 1) Quality standards

Importantly, standards defined the level at which a provider needs to function to reach certain outcomes, e.g. meeting minimum standards. Key mechanisms triggered by co-construction of setting standards included compliance, clarification, flexibility and adherence. Undergraduate and postgraduate QAPs responded to the regulators standards by inclusion in their own policies. The standards also provided QAPs with leverage to push forward changes at institutions:

> '*I think, a very positive element [O], is that it [standards] has allowed UK medical schools [C] within the framework to differ in how they implement that framework…So I think as well as having the rigour of what must be done, it allows for a degree of flexibility [M].' (HealthMedINT1)*
>
> *'We reference them [the standards] in our…internal policies.' (QAUG19)*

However, the standards had unintended consequences as their presence sometimes created confusion, particularly in the postgraduate context due to lack of clarity. In some instances, organisations had their own standards to assess educational quality, resulting in confusion:

> *'I suspect [postgraduate organisation] ignores them [standards] because they've come up with their own quality framework.' (QAPG33)*
>
> *'Our biggest concerns [O] really are not so much the standards as the sort of processes that by which the GMC [C] will check that our curricula are... comply with those standards [M]. And I think on that, you know, looking back now I can see that there was a certain ambiguity [M] in how the GMC were going to approach this and I'm not sure that they ever resolved it as the standards were being developed [O].' (QARC3)*

Standards that are overly prescriptive, rigid and inflexible prevent providers from being adaptable to need and innovation. For example standards which focus on particular aspects (e.g. student diversity) may detract attention from other areas of need (e.g. widening participation). Conversely, less binding standards (e.g. not detailing specific teaching methods) triggered mechanisms of ambiguity, openness and flexibility creating too much variation in education across contexts and producing new risks to quality. Misalignment between different quality standards caused frustration. For example, a LEP was deemed to be clinically outstanding but was also found to be inadequate for educational quality by a different regulator. Local pressures were seen to inhibit postgraduate partners' abilities to follow standards rigidly, suggesting that in the 'real world', applicability of standards was sometimes questionable:

> *'A lot of LEPs take our students [C], but they [LEPs] can quite readily tell us [medical school] to take them away as well [O], if we're very strict with them about meeting certain standards and certain criteria [M].' (QAUG5)*

Theme 2) Sanctions and approvals

We identified that organisational culture affects approaches to sanctions, and so an 'acceptable sanction' was contingent on risk. However in different contexts, should supportive measures be inadequate then the most severe sanction of withdrawn approval should remain. The "ultimate sanction of power" (QAPG28) was regarded to fulfil its intended consequence i.e. to protect patient safety, but also had unintended consequences to reinforce the medical regulator's authority and subsequently motivate providers to address problems. There was a firm belief that a severe sanction should rarely need to be enforced if other QA components (e.g. self-assessment) are effective.

> *'It's [sanctions] a bit of a lightning rod situation, but I think it [closing medical school] should remain as the ultimate sanction [O]…If trust management realised for example*

that they wouldn't lose their trainees as a result of not providing a safe and effective training environment [M]… I think [it would] slip further down their list of priorities [O].' (QARC24)

The effectiveness of regulatory approvals in the undergraduate and postgraduate context varied. In undergraduate, it was described as time-consuming examining both curriculum and staff capabilities. However, in the postgraduate context the process and need for QA was not as clear. Non-QAPs felt that it was important that mechanisms were in place to periodically review approvals. For instance:

'[The thoroughness] enabled them [regulator] to make a decision on our suitability to proceed.' (QAUG29)

'what if the trainee goes for one week, but it's only one week out of a one year placement, do they need to get that site approved?.' (QAPG12)

'We don't link approvals and quality very strongly [C]…we go to the GMC and we say, can we put some doctors here please? And the GMC go, yes. But there's an implication in doing that that because we're asking, we're going to quality manage that particular set of placements [M]. And we do, but not explicitly and not formally [O].' (QAPG16)

Theme 3) Collecting information: Visits, monitoring and self-assessment

Institutional visits were positioned as a key component as they triggered internal reviews and reflection, subsequently motivating organisations to improve quality. Working collaboratively engendered trust with open and honest dialogue, which was considered crucial in effecting change. Meaningful dialogue and collaboration were important and that was achieved through having high quality, "respected" (EducationINT32) trained visitor teams:

'I was prepared to be completely open and honest with the GMC…If [visits] are going to be effective, relationship building is actually more important than what you're doing collecting evidence.' (HealthMedUK21)

The QAF includes a range of monitoring data collection processes such as: data from the national training survey (NTS); monitoring including enhanced monitoring; self-assessment; and visits. The NTS surveys all doctors in training which facilitated increased transparency, accountability and risk-identification. The resources provided by the survey could lead to invaluable outcomes to examine training differences, make evidence-informed decisions and pinpoint training issues. However, multiple sources of data were sometimes regarded as conflictual, obscuring the overall picture of education quality.

'At the moment, they're [GMC] looking [at] trainee burnout [C]. So they're generating all this data at the moment and I don't think they're clear about what they're going to do with it [M], and my concern is they will just dump it on us for us to fix, and I don't think we can [O].' (QAPG10)

The component of requiring self-assessment triggered many different mechanisms in different contexts. For regulators, it generated reflective internally-led processes. The reasons for this were around connectivity between regulators and providers:

'[self-assessment] a really fundamental part of what we do, and we place a massive... emphasis on that.' (EducationUK17)

'this is the way [self-assessment] an institution connects itself with given standards" (EducationINT32)

'to work constructively with the provider…being the start of a peer review process' (HealthNon-MedINT11)

Whereas, for those who were being quality assured, the formality of written self-assessment inhibited open disclosure as it was laborious and seen more as an audit than self-assessment. Validity and reliability was also raised as perceptions of lack of feedback from the regulator also undermined self-assessment.

'I think you're more likely to hear genuine issues, genuine things that need to be fixed, if you speak to people informally and off the record' (QAPG6)

'It [self-assessment] forms part of it[assuring quality] and it's a very strong part of it, but I wouldn't necessarily use it [self-assessment] in isolation' (HealthMedUK23)

'*We are encouraging of institutions identifying challenges [C]. So if an institution is very open and honest [M], even into what might be quite a delicate area, saying this has been a challenge for us [C] and we're working away on it and we're doing the following things. Provided that their plan of action is a good one and that it's being conducted in a timely manner that would be reported on in a positive light [O].' (HealthNonMedUK9)*

Theme 4) Reporting; accountability, dissemination, good practice

A patient safety outcome response identified from external reporting was to build public confidence. Publicly available outputs fostered accountability to illustrate how providers are low-risk thus requiring less scrutiny. Insufficient reporting and feedback (in terms of timeliness, quantity and quality) fostered outcomes of devaluing time and effort, and subsequent disengagement. Risk context was also important to determine the effectiveness of intervention components. Informal partnerships were highlighted as critical to assuring quality.

'I think the transparency in publications are important because it involves or it makes things clear and open to all stakeholders.' (QAPG28)

'I think what having it public does, is it creates some pressure and accountability [M] on both the accreditor and the accrediting body [C] to focus on the outcomes and to show progress against conditions [O].' (HealthMedINT13)

'Working in partnership with regulators was instrumental, it has a significant effect on driving change in trusts' (HealthedUK21)

When feedback mechanisms triggered included collaboration and openness (rather than accountability) this fostered informal working partnerships leading to a raft of positive outcomes including awareness, sharing knowledge and quality improvement. Rapport over time helped provider's develop trust to report concerns. Here the positioning of regulator-provider context, moved from accountability quality-checker to collaborative problem-solver. Spreading good practice was contextually variable,

'The institution needs to take that genuine look at it, and spend the time genuinely evaluating and genuinely creating action plans…an institution that is good at critical self-reflection will tend to address problems before, or potential problems, before they become actual problems [O].' (EducationUK17)

'What works for one school may not work for other [O] …So, you don't want people to blindly be saying oh, let's do that, because that's going to be good practice here [M]….Because education in programs do vary [C]' (HealthNon-MedUK18)

Subsequently motivation reaction was low to implement changes based on other organisations' examples of good practice. Reports were deemed most helpful when they included action plans accessible to lay audiences.

DISCUSSION

Summary of findings

We undertook a realist evaluation to explore medical education QA undertaken by the GMC. We found that intervention components support or undermine QA for different organisations, and at different times in undergraduate (for medical students) and postgraduate (for trainees) contexts. We tested our initial programme theory (themes 1-4) to develop a modified programme theory (presented below) which articulates the need for components to be tailored proportionately to contexts.

Across the three-tier model we identified that the undergraduate and postgraduate context were influential. The leverage brokered by the regulator in the undergraduate context was often associated with directive features enabling local changes, whereas in contrast for the postgraduate context this power was often lost and diffused across education layers. Predominantly in the postgraduate trainee context, interventions led to unintended consequences (e.g. organisation disengagement) if an intervention promoted adherence at the cost of autonomy, subsequently triggering a lack of motivation (theme 1). An underlying mechanism identified to ensure an inclusive approach to QA was partnership working (themes 3-4). In the undergraduate context, provider-regulator engagement was sometimes not present, typically when there was a lack of informal relationships. In theme 3, visits were identified as a component that could better foster partnerships; so long as they were conducted with integrity, meaning and purpose.

Collectively across themes 1-4, the QAF was commended for being comprehensive and enabling a broad understanding of an organisation's performance. Internally-led processes with organisations identifying and addressing their own challenges and deficiencies, when done well, promoted a sense of autonomy and accountability (theme 3). The main unintended QAF consequences fell broadly into two outcomes, those related to the overlap across the three tiers (themes 1-2) and those related to the regulatory burden associated with data-driven approaches with a lack of transparency on why and how data was used (themes 3-4). A blurring of roles and boundaries of multiple organisations between patient safety, medical education and training was identified.

Modified programme theory

The findings informed a modified programme theory to explain the underlying processes for the intended and unintended consequences of how the GMC quality assure education in various contexts (Figure 4). Across the three-tier model (quality assurance, quality management and quality control), the theory demonstrates that QAF components are enacted differently with implications for the mechanisms triggered leading to positive (e.g. effecting change, contextual application of standards, partnerships) and/or negative outcomes (limited compliance, resistance, overlap). The influence and power of the regulator was continually picked up across the components (themes 1-4) which triggered mechanisms including transparent reporting to promote quality improvement, effective communication, trust, and partnership working facilitating interactions between regulators, partners and providers. Proportionate reactions in the face of disclosing and identifying patient safety risks at an early stage were more likely to occur within a positive trusting regulator-provider context underpinned by openness. Likewise, an organisation that self-assessed critically was reported to give regulators confidence in the institution.

Figure 4. Modified programme theory to explain the underlying process for the intended and unintended consequences of quality assurance components

[Insert figure 4 here]

Relevance of findings and implications

The findings reinforce the quality assurance literature highlighting *trust* in fostering effective working relationships to enhance feedback.[1 5] We extend this further, and identified that early communication of emerging risks supports quality assurance and enhancement approaches through informal networks. Visits aid communication and build relationships, yet if lost may distance the regulator and undermine opportunities for partnership working. Informal communication provided a safe environment for providers to discuss concerns with the regulator, opposed to formal monitoring acting as a barrier.

Expanding the literature, we demonstrate that context must be considered in order for quality assurance to protect patients.[23] Risk is context dependent, and was perceived to be tangibly different across undergraduate and postgraduate contexts. Undergraduate medical settings were perceived as low risk and imply opportunities for greater tailoring and focus. The overlaps between quality assurance, quality management and quality control were apparent especially within the postgraduate setting with duplication and confusion of responsibilities. These findings align with a recent systematic review identifying features of failing healthcare organisations including conflicting missions, fragmented accountability and lack of collaboration.[1]

Collectively, this supports the need to clarify structural quality processes and how organisations are intended to function collaboratively. In the analysis, risk-based visiting positioned the regulator as quality assurer rather than quality enhancer. Equally, effective assurance is often associated with suppressing innovation.[24] Moreover, the role of self-assessment[24] posed a number of challenges in relation to purpose and autonomy. While institutional self-assessments can positively influence reactions to drive quality improvement, there are issues with validity, reliability and internal quality review.[19 24-27]

The power of the regulator impacted on the effectiveness of intervention components in multiple ways. The regulatory-burden associated with monitoring activities was considerable and disengagement ensued. Lack of feedback from the regulator was an important aetiological mechanism precipitating the situation. Similarly, negative consequences of approvals including cost, low staff morale, threats to organisation reputation, and the suppression of innovation through adhering to standards has been identified.[28 29] Without regulators addressing varying risk contexts, the proportionality of QA is imbalanced, leading to negative outcomes with regulators unable to effectively assure quality. Therefore, collectively considering a hybrid model of cyclical plus risk-based visiting may help to build provider relationships and drive improvement whilst ensuring minimum standards. Collective assurance and relationships should be encouraged so that regulators and providers can tackle issues conjointly. Flexibility in utilising other datasets within any collaborative work is a necessity and a clear stance on organisational remit and particularly boundaries, is anticipated to be a key mechanism in effective joint QA.

Strengths, limitations and future directions

To our knowledge this is the first robust study on education QA within the healthcare context, synthesising data from stakeholders. The study fills a gap as QA remains expensive, yet its functionality is largely unexplored. The study was conducted by an experienced multidisciplinary research team applying an innovative realist approach, underpinned by a sound team-based analysis. A somewhat surprising omission from our findings is a lack of attention to the mechanism of leadership.[1 23] The sample focussed on processes rather than delivery perhaps contributing to such omission. Moving forward, there is a need to conduct an economic review and consult with stakeholders into what data could be shared (e.g. National Training Survey, Care Quality Commission data) to understand links to intervention components. The

findings have influenced the GMC's approach to QA which impacts on all medical students and doctors in training.[30]

Conclusions

This study used a realist methodology to reveal the intended and unintended consequences of components used by the GMC to quality assure medical education, and elucidated the mechanisms by which both are brought about. While uniform approaches are often in place, interventions need to be contextually tailored. Ongoing partnership working can enhance open disclosure to drive up education quality. This research has provided a modified programme theory to explicate how education providers and regulators can work more effectively together to uphold quality, and ultimately protect public safety.

Acknowledgements

We would like to thank the participants for sharing their insights, the GMC for facilitating recruitment, and Dr Graham Easton and UCL Design team for helping to produce the informational video.

Contributorship

AG, PC, LM and MP conceived the study design. PC and LM designed study materials and gained ethical approval. All authors collected data, contributed to data analysis. PC wrote the first draft of the manuscript with input from AG, MP, LK and LM. All authors approve the final manuscript.

Data sharing

We do not have ethical permission to share raw data.

Funding and conflict of interest

This study was commissioned by the GMC, project ID: GMC 822. AG was the principal investigator.

Ethics

This study was registered with UCL's data protection office on 06/06/18 and approved by UCL ethics on 15/06/18, project ID: 6281/003. All participants volunteered to take part and provided written consent. Data were anonymised with respect to individuals and institutions.

References

1. Vaughn VM, Saint S, Krein SL, et al. Characteristics of healthcare organisations struggling to improve quality: results from a systematic review of qualitative studies. *BMJ Qual Saf 2019*, *28*(1), pp.74-84.

2. General Medical Council. Promoting excellence 2019. Accessed on 16.07.19 from: https://www.gmc-uk.org/education/standards-guidance-and-curricula/standards-and-outcomes/promoting-excellence

3. Vroeijenstijn, A.I. Quality assurance in medical education. *Academic medicine: journal of the Association of American Medical Colleges* 1995, *70*(7 Suppl), pp.S59-67.

4. General Medical Council (2019). How we assure quality. Accessed on 10/04/2019 from: https://www.gmc-uk.org/education/how-we-quality-assure

5. Jones K & Tymms P. Ofsted's role in promoting school improvement: the mechanisms of the school inspection system in England, *Oxford Review of Education 2014*, 40(3), 315-330, DOI: 10.1080/03054985.2014.911726

6. Westerheijden DF, Stensaker B, Rosa MJ (eds). *Quality assurance in higher education: Trends in regulation, translation and transformation* (Vol. 20) 2007. Springer Science & Business Media.

7. Haji F, Morin MP, Parker K. Rethinking programme evaluation in health professions education: beyond 'did it work?'. *Medical education 2013*, *47*(4), pp.342-351.

8. Øvretveit J, Gustafson D. Evaluation of quality improvement programmes. *BMJ Qual Saf* 2002;11:270-275.

9. Pawson R, Greenhalgh T, Harvey G, et al. Realist review – a new method of systematic review designed for complex policy interventions, *Journal of Health Services Research & Policy 2005* 10(1): 21–34

10. Pawson R. *The Science of Evaluation: A Realist Manifesto*, SAGE Publications Ltd, London, 2013.

11. Jones AC, Shipman SA, Ogrinc G. Key characteristics of successful quality improvement curricula in physician education: a realist review BMJ Qual Saf 2015;24:77-88

12. Wong G, Westhorp G, Manzano A, et al. RAMESES II reporting standards for realist evaluations. *BMC medicine 2016*, *14*(1), p.96.

13. Griffin A, Crampton P, Mehdizadeh L, et al. Understanding stakeholder perspectives on the GMC's quality assurance of medical education and training. Final Report, 2018. Accessed 25.07.19 from: https://www.gmc-uk.org/-/media/documents/understanding-stakeholder-perspectives-on-the-gmcs-quality-assurance-of-medical-education-and-traini.pdf?la=en&hash=14529235D9E6C99AF9B5EDA82F6848CB34D33396

14. Jagosh J, Bush PL, Salsberg J et al. A realist evaluation of community-based participatory research: partnership synergy, trust building and related ripple effects. *BMC public health 2015*, *15*(1), p.725.

15. Manzano A. The craft of interviewing in realist evaluation. Evaluation. 2016 Jul;22(3):342-60.

16. Keren G, De Bruin WB. On the assessment of decision quality: Considerations regarding utility, conflict and accountability. Thinking: Psychological perspectives on reasoning, judgment and decision making. 2003 Sep 16;347.

17. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. *The qualitative researcher's companion* 2002; 573, 305-329.

18. NVivo qualitative data analysis software; QSR International Pty Ltd. Version 11, 2015.

19. Griffin A, McKeown A, Viney R, et al. Revalidation and quality assurance: the application of the MUSIQ framework in independent verification visits to healthcare organisations. *BMJ Open* 2017;7:e014121. doi:10.1136/bmjopen-2016-014121

20. Griffin A, Knight L, McKeown A, et al. A postgraduate curriculum for integrated care: a qualitative exploration of trainee paediatricians and general practitioners' experiences. *BMC Med Educ* 2019; *19*(1), p.8.

21. Rees CE, Crampton PES, Kent F, et al. Understanding students' and clinicians' experiences of informal interprofessional workplace learning: an Australian qualitative study. BMJ open 2018.1;8(4):e021238.

22. Ajjawi R, Crampton PES, Rees CE. What really matters for successful research environments? A realist synthesis. *Medical education 2018*, *52*(9), pp.936-950.

23. Kaplan HC, Provost LP, Froehle CM, & Margolis PA. The Model for Understanding Success in Quality (MUSIQ): building a theory of context in healthcare quality improvement. *BMJ Qual Saf 2012*, *21*(1), pp.13-20.

24. Lang DW. Self-regulation with rules: Lessons learned from a new quality assurance process for Ontario, *Quality Assurance in Education 2015*, 23:3, 216-232, https://doi.org/10.1108/QAE-09-2014-0046

25. Tackett S, Grant J & Mmari K. Designing an evaluation framework for WFME basic standards for medical education, *Medical Teacher 2016*, 38(3), 291-296, DOI:10.3109/0142159X.2015.1031737

26. Jones KL, Tymms P, Kemethofer D, et al. The unintended consequences of school inspection: the prevalence of inspection side-effects in Austria, the Czech Republic, England, Ireland, the Netherlands, Sweden, and Switzerland, *Oxford Review of Education 2017*, 43(6), 805-822, DOI: 10.1080/03054985.2017.1352499

27. Forrest C. Inspection and improvement in three further education colleges, *Research in Post-Compulsory Education 2015*, 20(3), 296-314, DOI: 10.1080/13596748.2015.1063273

28. de Paor C. The contribution of professional accreditation to quality assurance in higher education, *Quality in Higher Education 2016*, 22(3), 228-241, DOI: 10.1080/13538322.2016.1263925

29. Blouin D, Tekian A, Kamin C, et al. The impact of accreditation on medical schools' processes, *Medical Education* 2018, 52(2), 182-191

30. General Medical Council. Council Papers February 2019. Accessed on 17.07.19 from: https://www.gmc-uk.org/-/media/documents/council---agenda-and-papers-for-meeting-on-27-february-2019_pdf-78924314.pdf

Figure 1: The intervention: quality assurance framework (QAF)

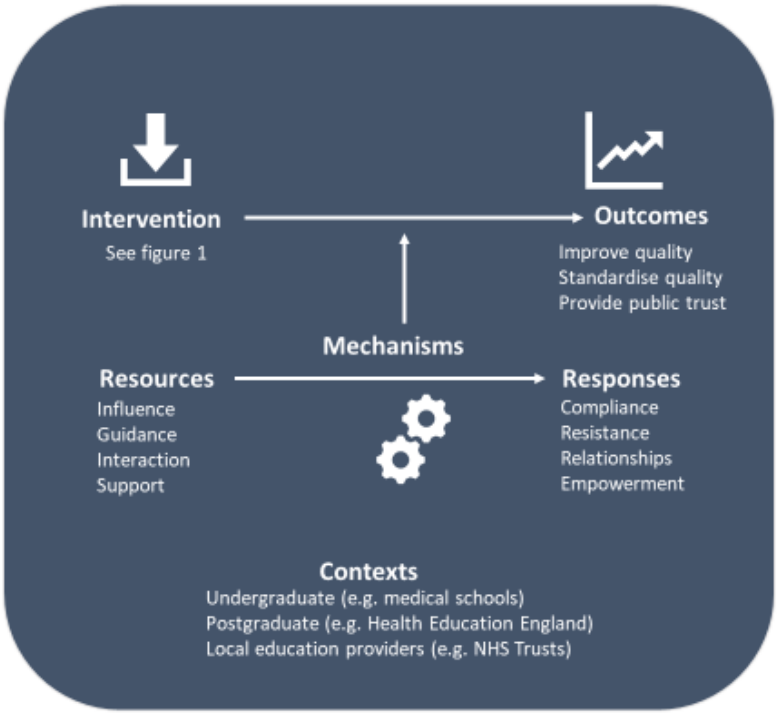


Figure 2. Initial programme theory

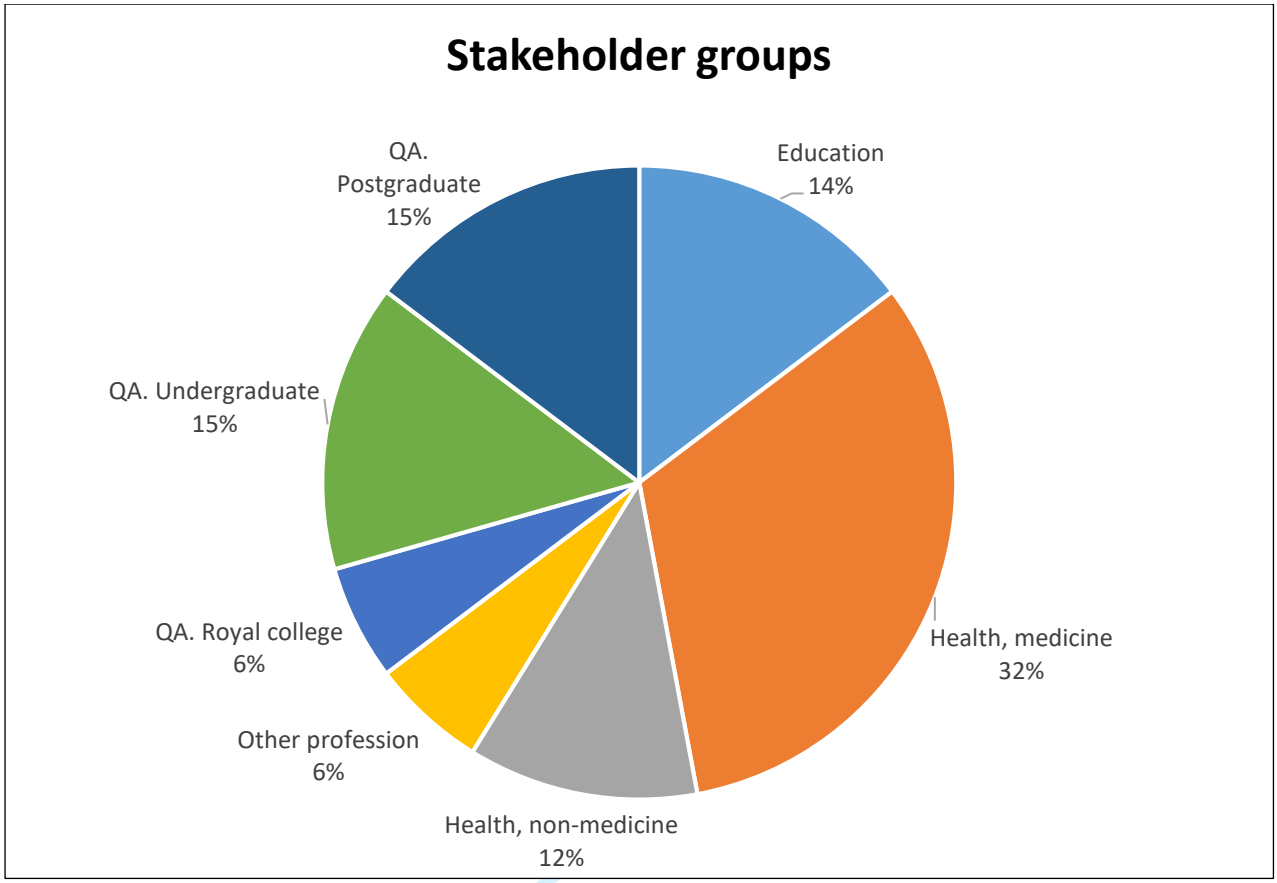


Figure 3. Stakeholder groups

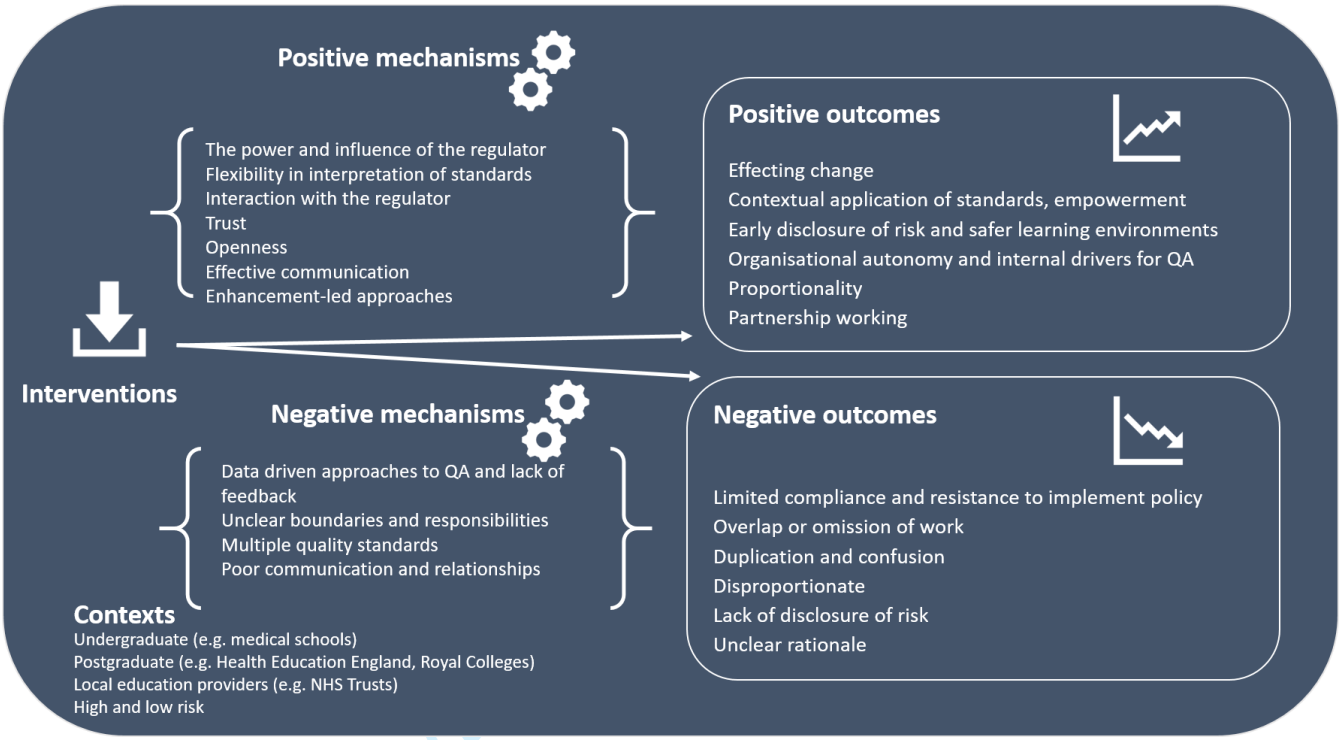


Figure 4. Modified programme theory to explain the underlying process for the intended and unintended consequences of quality assurance components

1. Quality Assurance partners interview schedule

Structure

Introduce and duration about 45 mins

Process

This interview will be recorded and analysed, looking for common themes that arise across the interviews.

Consent

A reminder that we will not personally name anyone who takes part in the study. Do you have any questions before we start the tape? Thank you for signing the consent form [or take verbal consent if required].

Introductory/ Background questions

- How long have you worked in your quality assurance role?
- Who else in your organisation works in a QA role?

General: GMC quality assurance framework overall

- Are there any aspects of GMC's quality assurance framework that you think are particularly effective, i.e. give you reassurance in their processes?
- Are there any aspects of the framework that you think are less effective or problematic in some way, i.e. do not assure you?

Focused: Specific aspects of GMC's quality assurance framework

I would now like to ask some questions about different components of the QA framework and be keen for you to share your experiences where relevant.

Standards

- Are the standards the right ones? Prompt any missing?
- Are the standards helpful or unhelpful in anyway?
- Has using the standards had any impact on your organisation?

Approvals

- What would be the advantages of making the GMC's approvals time limited?
- What would be the disadvantages of making the date GMC's approvals time limited?
- Do you think the GMC's approvals process is effective?

Monitoring:

- Is the GMC monitoring the right evidence to assess your organisations performance?

- What sources of evidence do you think give the GMC the best insight into your organisation?
- What other areas could they/should they monitor?
- Does monitoring have any impact on your organisation? Prompt: Positive/negative
- Turning to enhanced monitoring, some people would say that the GMC are overstepping their remit when they require postgraduate organisations to report training programs and local education providers to them, what are your thoughts?

Sharing evidence:

- How effective is the GMC at sharing evidence with you?
- Is there evidence that could be shared more effectively and how would that benefit your organisation?
- Is there any evidence that you feel should not be shared, in particular with other healthcare regulators?

Self-assessment:

The GMC requires annual self-assessment from the medical Royal colleges and medical schools but not the postgraduate organisations.

- Do you think self-assessment is a helpful process?
- Some hold the view that organisational self-assessment is not a reliable process, what do you think?
- Has the process of self-assessment resulted in any organisational change?

Visits:

- What purpose do you think visits to organisations have?
    - Prompt: What makes a visit effective?
    - Prompt: What are the important areas that visits should include?
- Most regulators are moving away from cyclical or scheduled visiting, towards entirely risk-based systems, however many GMC stakeholders believe the cyclical visits have many benefits and should be retained. What do you think?
- What would happen if the GMC did no visiting?

Reporting:

The GMC currently publishes long-form visit reports on its website, as well as information about enhanced monitoring cases, and data tools such as the NTS reporting tool and the progression reports.

- What do you think of the current QA reporting?
- Are there any negative consequences of reporting data on the website?
- Does your organisation use the reports in anyway?

Good practice:

The GMC aim to identify good practice across medical schools and postgraduate bodies and then publish this on their website.

- Is this useful to your organisation?
    - Prompt: positive aspects v negative

- Have you adopted any areas of good practice yourself?
- Some people would say more resources should go into quality enhancement rather than accountability. What are your views?

Fairness

- How can the GMC quality assure fairness in medical education and training?

Sanctions:

Sanctions mean withdrawing trainee doctors from the NHS or closing down medical schools which has a critical impact on healthcare.

- In the case of an underperforming training organisation that is currently failing to meet required standards what might be a proportionate sanction from the GMC that is not as extreme as withdrawing approval?
    - Prompts: The GMC visiting, publicly available rating scales, time bound approvals

Collective assurance

The GMC has committed to working more closely with other regulators to find efficiencies and reduce the regulatory burden on the service.

- What would be the advantages for your organisation in this approach?
- Would there be any disadvantages?
- Would sharing data enable the GMC to identify risk better?
- How practical would it be for your organisation to undertake joint visiting?
- Do you think the GMC's approach to QA is proportionate to the risks involved in medical education and training?
- Do you have any suggestions for improving the GMC's quality assurance processes?

Thank you for your time. Is there anything you would like to add that we haven't discussed?

Thank you

2. Quality Assurance non-partners interview schedule

Structure

Introduce and duration about 45 mins

Process

This interview will be recorded and analysed, looking for common themes that arise across the interviews.

Consent

A reminder that we will not personally name anyone who takes part in the study. Do you have any questions before we start the tape? Thank you for signing the consent form [or take verbal consent if required].

Introductory/ Background questions

- Can you briefly explain the context in which your organisation is involved in QA
- What is your specific role?
- How long have you worked in your quality assurance role?
- Who else in your organisation works in a QA role?

General: GMC quality assurance framework overall

- Are there any aspects of GMC's quality assurance framework that you think are particularly effective, i.e. give you reassurance in their processes?
- Are there any aspects of the framework that you think are less effective or problematic in some way, i.e. do not assure you?

Focused: Specific aspects of GMC's quality assurance framework

I would now like to ask some questions about different components of the QA framework and be keen for you to share your experiences where relevant.

Standards

- Are the standards the right ones?
    - Prompt: Any missing?
- Are the standards helpful or unhelpful in anyway?

Approvals

- Do you think the GMC's approvals process is effective?
- What would be the advantages of making the GMC's approvals time limited?
- What would be the disadvantages of making the date GMC's approvals time limited?

Monitoring

- Is the GMC monitoring the right evidence to assess organisational performance?
    - Prompt: What other areas could they/should they monitor?
- Do you think monitoring has any impact on organisation performance?
    - Prompt: Positive/negative
- Turning to enhanced monitoring, some people would say that the GMC are overstepping their remit when they require postgraduate organisations to report training programs and local education providers to them, what are your thoughts?
- How does your organisation use monitoring?
- Do you have a model for triangulating predicting risk?

Sharing evidence

- How could the GMC improve sharing its evidence?
    - Prompt: Between regulator to regulator; between regulators to QA partners?
- Is there other evidence that could be shared?
    - Prompt: Is there any evidence that you feel should not be shared?

Self-assessment

The GMC requires annual self-assessment from the medical Royal colleges and medical schools but not the postgraduate organisations.

- Do you think self-assessment is a helpful process?
- Some hold the view that self-assessment is not a reliable process, what do you think?
- What is your organisations approach to self-assessment?
- Has the process of self-assessment resulted in any organisational change?

Visits

- What purpose do you think visits to organisations have?
    - Prompt: What makes a visit effective?
    - Prompt: What impact do you think organisational visits have?
- Most regulators are moving away from cyclical or scheduled visiting, towards entirely risk-based systems, however many GMC stakeholders believe the cyclical visits have many benefits and should be retained. What do you think?
- How can visits give greater assurance of quality?
- What would happen if the GMC did no visiting?

Reporting

The GMC currently publishes long-form visit reports on its website, as well as information about enhanced monitoring cases, and data tools such as the NTS reporting tool and the progression reports.

- What do you think of the GMC approach to reporting?
- How does your organisation report on performance?
    - Prompts: strengths and weaknesses?

Good practice

The GMC aim to identify good practice across medical schools and postgraduate bodies and then publish this on their website.

- What do you think of the GMCs approach to sharing best practice?
- Some people would say more resources should go into quality enhancement rather than accountability. What are your views?
- What is your organisations approach to this?

Fairness

- How can the GMC quality assure fairness in medical education and training?

Sanctions

In the GMC's context, sanctions mean withdrawing trainee doctors from the NHS or closing down medical schools which has a critical impact on healthcare.

- Does your organisation have any advice or experience of imposing meaningful sanctions that would not be considered as extreme as the GMC's approach?

Collective assurance

The GMC has committed to working more closely with other regulators to find efficiencies and reduce the regulatory burden on the service.

- Is your organisation involved in joint visits? If so, what would be the advantages/ disadvantages for your organisation in this approach?
- Do you sharing data with other organisations?
- How practical would it be for your organisation to undertake joint visiting?
- Do you think the GMC's approach to QA is proportionate to the risks involved in medical education and training?
- Do you have any suggestions for improving the GMC's quality assurance processes?

Thank you for your time. Is there anything you would like to add that we haven't discussed?

Thank you

RAMESES 2 reporting checklist

Source: Wong, G., Westhorp, G., Manzano, A., Greenhalgh, J., Jagosh, J. and Greenhalgh, T., 2016. RAMESES II reporting standards for realist evaluations. *BMC medicine*, *14*(1), p.96.

TITLE			**Reported in document**	**Section, Page(s) in document**
1		In the title, identify the document as a realist evaluation	Y	Title, pg 1
SUMMARY OR ABSTRACT				
2		The abstract or summary should include brief details on: the policy, programme or initiative under evaluation; programme setting; purpose of the evaluation; evaluation question(s) and/or objective(s); evaluation strategy; data collection, documentation and analysis methods; key findings and conclusions Where journals require it and the nature of the study is appropriate, brief details of respondents to the evaluation and recruitment and sampling processes may also be included Sufficient detail should be provided to identify that a realist approach was used and that realist programme theory was developed and/or refined	Y	Abstract, pg 2
INTRODUCTION				
3	Rationale for evaluation	Explain the purpose of the evaluation and the implications	Y	Introduction pg. 3-4

TITLE			Reported in document	Section, Page(s) in document
		for its focus and design		
4	Programme theory	Describe the initial programme theory (or theories) that underpin the programme, policy or initiative	Y	Methodology pg 4
5	Evaluation questions, objectives and focus	State the evaluation question(s) and specify the objectives for the evaluation. Describe whether and how the programme theory was used to define the scope and focus of the evaluation	Y	Introduction, study aim, pg 3-5
6	Ethical approval	State whether the realist evaluation required and has gained ethical approval from the relevant authorities, providing details as appropriate.	Y	Ethical approval, pg 5 All participants volunteered to take part and provided written consent.
METHODS				
7	Rationale for using realist evaluation	Explain why a realist evaluation approach was chosen and (if relevant) adapted	Y	Methods, conceptual framework, pg 4-5
8	Environment surrounding the evaluation	Describe the environment in which the evaluation took place	Y	Methods, conceptual framework, data collection, pg 4-6
9	Describe the programme policy, initiative or product evaluated	Provide relevant details on the programme, policy or initiative evaluated	Y	Introduction, pg 3-4; Methods, conceptual framework, Pg 5-6
10	Describe and	A description and justification	Y	Introduction, pg

TITLE			Reported in document	Section, Page(s) in document
	justify the evaluation design	of the evaluation design (i.e. the account of what was planned, done and why) should be included, at least in summary form or as an appendix, in the document which presents the main findings. If this is not done, the omission should be justified and a reference or link to the evaluation design given. It may also be useful to publish or make freely available (e.g. online on a website) any original evaluation design document or protocol, where they exist		4; Methods, conceptual framework, Pg 5-6
11	Data collection methods	Describe and justify the data collection methods – which ones were used, why and how they fed into developing, supporting, refuting or refining programme theory Provide details of the steps taken to enhance the trustworthiness of data collection and documentation	Y	Methods, data collection, pg 4-5
12	Recruitment process and sampling strategy	Describe how respondents to the evaluation were recruited or engaged and how the sample contributed to the development, support, refutation or refinement of programme theory	Y	Methods, sampling, recruitment, pg 4-5
13	Data analysis	Describe in detail how data	Y	Methods,

TITLE			Reported in document	Section, Page(s) in document
		were analysed. This section should include information on the constructs that were identified, the process of analysis, how the programme theory was further developed, supported, refuted and refined, and (where relevant) how analysis changed as the evaluation unfolded		analysis, reflexivity, pg 5
RESULTS				
14	Details of participants	Report (if applicable) who took part in the evaluation, the details of the data they provided and how the data was used to develop, support, refute or refine programme theory	Y	Results, participant details, pg 5
15	Main findings	Present the key findings, linking them to contexts, mechanisms and outcome configurations. Show how they were used to further develop, test or refine the programme theory	Y	Results, main findings, pg 5-9
DISCUSSION				
16	Summary of findings	Summarise the main findings with attention to the evaluation questions, purpose of the evaluation, programme theory and intended audience	Y	Discussion, summary of key findings, pg 9-11
17	Strengths, limitations and future directions	Discuss both the strengths of the evaluation and its limitations. These should include (but need not be limited	Y	Discussion, Strengths, limitations and future directions, pg

TITLE			Reported in document	Section, Page(s) in document
		to): (1) consideration of all the steps in the evaluation processes; and (2) comment on the adequacy, trustworthiness and value of the explanatory insights which emerged The particular implications arising from the realist nature of the findings should be reflected in these discussions		11-12
18	Comparison with existing literature	Where appropriate, compare and contrast the evaluation's findings with the existing literature on similar programmes, policies or initiatives	Y	Discussion, Comparison with existing literature and implications, pg 10-12
19	Conclusion and recommendations	List the main conclusions that are justified by the analyses of the data. If appropriate, offer recommendations consistent with a realist approach	Y	Discussion, implications, conclusions, pg 10-12
20	Funding and conflict of interest	State the funding source (if any) for the evaluation, the role played by the funder (if any) and any conflicts of interests of the evaluators	Y	Funding and competing interests, pg 12 This study was externally commissioned by the GMC

BMJ Open

A realist evaluation of UK medical education quality assurance

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-033614.R2
Article Type:	Original research
Date Submitted by the Author:	27-Nov-2019
Complete List of Authors:	Crampton, Paul; Hull York Medical School, Health Professions Education Unit; UCL, Research Department for Medical Education Mehdizadeh, Leila; UCL, Research Department for Medical Education Page, Michael; QMUL, Institute for Health Sciences Education Knight, Laura; UCL, Research Department for Medical Education Griffin, Ann; University College London, Medical School
Primary Subject Heading:	Medical education and training
Secondary Subject Heading:	Medical education and training, Qualitative research
Keywords:	QUALITATIVE RESEARCH, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, MEDICAL EDUCATION & TRAINING

Title page

Title: A realist evaluation of UK medical education quality assurance

Authors: Paul E.S. Crampton[1,3] , Leila Mehdizadeh[1] , Michael Page[4] , Laura Knight[1] , Ann Griffin[2]

Address affiliations:

[1] Research Department of Medical Education, UCL Medical School, Royal Free Hospital, Room GF/664, London, NW3 2PF, UK

[2] Research Department of Medical Education, UCL Medical School, The Directorate, 74 Huntley Street, London, WC1E 6AU, UK

[3] Hull York Medical School, York University, John Hughlings Jackson Building, University Rd, Heslington, York YO10 5DD, UK

[4] Institute for Health Sciences Education, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, Room 3.15, 3rd Floor, Garrod Building, Turner Street, Whitechapel, London, E1 2AD UK

Corresponding author: Ann Griffin, email: a.griffin@ucl.ac.uk, Research Department of Medical Education, UCL Medical School, The Directorate, 74 Huntley Street, London, WC1E 6AU, UK

Word count: 4224

Key words (MESH): Quality assurance, health care; Qualitative research; Medical Education

ABSTRACT

Objectives: The aim of the study was to explore what components of the General Medical Council's (GMC) quality assurance framework work, for whom, in what circumstances and how?

Setting: UK undergraduate and postgraduate medical education and training.

Participants: We conducted interviews with a stratified sample of 36 individuals. This included those who had direct experiences, as well as those with external insights, representing local, national and international organisations within and outside medicine.

Intervention: The GMC quality assure education to protect patient and public safety utilising complex intervention components including meeting standards, institutional visits and monitoring performance. However, the context in which these are implemented matters. We undertook an innovative realist evaluation to test an initial programme theory. Data were analysed using framework analysis.

Results: Across components of the intervention, we identified key mechanisms including: transparent reporting to promote quality improvement; dialogic feedback; partnership working facilitating interactions between regulators and providers, and; role clarity in conducting proportionate interventions appropriate to risk. The GMC's framework was commended for being comprehensive and enabling a broad understanding of an organisation's performance. Unintended consequences included confusion over roles and boundaries in different contexts which often undermined effectiveness.

Conclusions: This realist evaluation substantiates the literature and reveals deeper understandings about quality assuring medical education. While standardised approaches are implemented, interventions need to be contextually proportionate. Routine communication is beneficial to verify data, share concerns and check-risk; however, ongoing partnership working can foster assurance. The study provides a modified programme theory to explicate how education providers and regulators can work more effectively together to uphold education quality, and ultimately protect public safety. The findings have influenced the GMC's approach to quality assurance which impacts on all medical students and doctors in training.

ARTICLE SUMMARY

Strengths and limitations of this study

• The study is underpinned by a sound team-based reflexive analysis and is the first in-depth realist evaluation of quality assurance in the healthcare context

• The study fills a critical knowledge gap on quality assurance and the findings have influenced the General Medical Council's approach which impacts on all medical students and doctors in training

• To enhance the transferability of our findings, we collected qualitative data from a large number of UK and international based expert stakeholders within and outside medicine

• A limitation is that the study is specific in its UK focus and focuses on the General Medical Council

• Furthermore the study did not analyse the impact of outcome data (e.g. National Training Survey, Care Quality Commission data) to understand links between quality assurance and intervention components

INTRODUCTION

Rationale

Healthcare regulators quality assure education and protect public safety utilising complex intervention components including setting standards, institutional visits and monitoring performance. However, the context in which these components are implemented matters.[1-3] Within undergraduate and postgraduate medical education and training, the taught curriculum integrates with workplace-based experiential learning. Consequently education environments range from structured classrooms in university contexts to clinical placements within shifting healthcare structures across primary and secondary care. Therefore, the challenge for regulators is to mediate the quality of education and training across these spaces in order to assure the public that education and training is safe, that all medical students are prepared for practice and that all trainee doctors are fit to practise.

In the UK context, the General Medical Council (GMC) regulator work closely with other organisations to secure its standards, using a three-tier model. The GMC (Tier 1, quality assurance), has an overarching responsibility to hold undergraduate and postgraduate training bodies to account for meeting standards. These bodies (Tier 2, quality management) organise, manage, commission, and sometimes deliver medical education. They also manage quality in local education providers (LEPs), where students and trainees are placed, such as trusts, health boards, general practices, and other clinical settings. The LEPs (Tier 3, quality control), have processes to ensure satisfactory clinical placements, and that their organisation provides an appropriate learning environment. Medical royal colleges work with the GMC to ensure their curricula and assessments are fit for purpose, inform specialty and postgraduate programme delivery, and have local systems to support training.

The GMC has a multifaceted intervention to examine the quality of medical education and training provision, known as the quality assurance framework (QAF, figure 1). The intervention includes the following components: setting standards, approving education settings, monitoring activities including self-assessment and enhanced monitoring, visits, sharing evidence with other regulators and identifying good practice.[4] The QAF operates across the three-tier model i.e. between the regulator (e.g. GMC), organisational provider bodies (e.g. medical schools) and local service delivery organisations (e.g. hospitals, general practice).

Figure 1: The intervention: quality assurance framework (QAF)

[Insert figure 1 here]

A range of approaches can be implemented to assure education quality, from heavily arbitrated measures to informal uncontrolled processes. Existing education Quality Assurance (QA) research is sparse and tends to come from the field of school-based and higher education.[5 6] Whilst exploring the mechanisms of action of school inspections,[5] a theory stated that regulatory activities associated with improvement include: setting standards; the provision of feedback; employing a system of sanctions and rewards; monitoring schools by the collection of information and public accountability. However, more research is needed to understand within the healthcare context how quality assurance can protect patients. Despite large amounts of resources dedicated to education QA, there remains a lack of clear evidence. QA takes place within varied and complex social environments. For this reason, the same intervention can impact on individuals, teams and organisations in different ways.[7 8] Although there are intended consequences explicit in the QAF design, the implicit underlying drivers of these are not clear.

Specific aim

The study aim was to explore what components of the GMC's quality assurance framework work, for whom, in what circumstances and how?

METHODOLOGY

Conceptual framework: realist evaluation

We chose realist methodology because of a focus on four theoretically constructed and inter-related questions: what works, for whom, in what circumstances, and how?[9-11] This results in generative causation, about how QAF components operate, offering an assessment of *whether* they work, as well as *why*. In the results, we explore the complex configuration links between contexts (where, when and with whom the activity takes place), components (different activities applied to assure quality), mechanisms (underlying processes for why the activity is/is not effective), and outcomes (intended and unintended consequences).[12] Our methods follow the RAMESES 2 (realist and meta-narrative evidence synthesis: evolving standards)[12] reporting guidelines (for full report see Griffin).[13]

Initial programme theory

We developed an initial programme theory (figure 2) based on existing literature, the GMC's approach to QA and research team insight (see reflexivity). We positioned the QAF as consisting of various components which we then explored to answer our study aim. Each component triggers multiple responses when applied in certain contexts with underpinning resources. Our programme theory postulated that within undergraduate, postgraduate and local education provider contexts, the QAF led to improved quality and protected the public through exploiting regulatory influence, guidance and supporting organisations, leading to compliance, resistance, relationships and empowerment.

Figure 2. Initial programme theory

[Insert figure 2 here]

Sampling and recruitment

Stratified purposive sampling was used to test our theory with stakeholders. We targeted those who were familiar with the QAF, labelled in the study as quality assurance partners (QAPs); as well as those with outsider perceptions of how the framework is positioned in society (e.g. other regulators) and broader regulation contexts (e.g. education). These spanned organisations both inside and outside healthcare, including internationally; collectively labelled as non-quality assurance partners (non-QAPs). See table 1 for the list of QAPs.

Our realist position acknowledged that stakeholders each had partial knowledge of the intervention, therefore to fully explore research questions we included policy makers, implementers and recipients.[14] Following email invitations, non-responders received two reminders. Aligned to realist evaluation,[15] participants were given the opportunity to review project materials prior to participation via a 15-minute informational video.

Data collection

To test our programme theory we undertook semi-structured interviews to explore underlying processes triggered by QA components (Supplementary file 1). We tested a number of candidate theories to explore the underlying ways in which the intervention was intended to be successful. For example, to test theories on the impact of generalisability and utility[16] we asked: "Some hold the view that organisational self-assessment is not a reliable process, what do you think?", and "What would happen if the medical regulator (GMC) did no organisational visiting?" We designed questions with different foci appropriate to participants,[15] for QAPs and non-QAPs. Additionally, researchers probed for reasoning when participants

gave limited responses. Interview questions were piloted with a QA manager within our own medical school. Very minor changes were required to enhance clarity, appropriateness and sense-making. All interviews were conducted one-to-one apart from two, where two people from the same organisation were present to provide comprehensive responses. Telephone/skype interviews were conducted by four members of the team (AG, PC, LM, MP), audio-recorded, and transcribed verbatim. Participants were geographically dispersed across the world therefore face-to-face interviews were not feasible.

Data analysis

Data were analysed following the stages of framework analysis.[17] This approach was followed due to its retroductive inductive-deductive nature to test initial theory whilst identifying emerging Context, Mechanism and Outcomes (CMOs) configurations. For *familiarisation*, two researchers each read one transcript and then made notes to identify CMO configurations (Total: 5 researchers X 5 transcripts). All researchers met to discuss similarities and differences across the transcripts including recurrent CMO configurations. We then developed a coding framework, consistently applied to each transcript. Five researchers coded data using NVivo,[18] with frequent progress meetings.

Reflexivity

Prior to analysis, individual members completed a written reflexive exercise which highlighted prior dispositions towards the research, and were then discussed collectively. The team consisted of practitioners, academics and researchers from medical and social science background disciplines, with QA knowledge ranging novice-expert.[19] Team members were vastly experienced in qualitative methods[20 21] and had previously applied a realist lens to understand complex interventions in healthcare education.[22]

Patient and public involvement

Given the focus on regulator-medical provider interactions in this study, patients and the public were not involved in the design, data collection or data analysis.

RESULTS

Participant details

Following ethical approval, we conducted interviews with 36 individuals representing 34 different organisations between July-September 2018 (table 1) to produce a considerable amount of data: 35 hours, 27 minutes. Interviews ranged between 48-88 minutes, mean = 63. The sample represented regional, national (England, Scotland, Northern Ireland, Wales) and international stakeholders within and outside medicine (Figure 3). There were 12 QAPs and 22 non-QAPs, with 27 (79%) of these from the UK and 7 (21%) international, representing Asia, North America and Europe. Participants often held senior roles such as chief executives and quality leads.

Table 1. Demographic information of the participant organisations

Key: QA=Quality Assurance partner, Health=health organisation, Med=medicine organisation, UG=undergraduate, PG=postgraduate, RC=royal college, INT=international based, UK=United Kingdom based

Interview number	ID	QA partner (Y/N)	Profession Sector	Location (coverage)
1	HealthMedINT1	N	Health, medicine	International
2	QAUG2	Y	Undergrad	UK

3	QARC3	Y	Royal College	UK
4	OtherprofessionUK4	N	Other Profession	UK
5	QAUG5	Y	Undergrad	UK
6	HealthMedUK6	N	Health, medicine	UK
7	HealthMedUK7	N	Health, medicine	UK
8	EducationUK8	N	Education	UK
9	HealthNon-MedUK9	N	Health, non-medicine	UK
10	QAPG10	Y	Postgrad	UK
11	HealthNon-MedINT11	N	Health, non-medicine	International
12	QAPG12	Y	Postgrad	UK
13	HealthMedINT13	N	Health, medicine	International
14	QAPG14	N	Health, medicine	UK
15	QAUG15	Y	Undergrad	UK
16	QAPG16	Y	Postgrad	UK
17	EducationUK17	N	Education	UK
18	HealthNon-MedUK18	N	Health, non-medicine	UK
19	QAUG19	Y	Undergrad	UK
20	HealthMedINT20	N	Health, medicine	International
21	HealthMedUK21	N	Health, medicine	UK
22	OtherprofessionUK22	N	Other profession	UK
23	HealthMedUK23	N	Health, medicine	UK
24	QARC24	Y	Royal College	UK
25	EducationUK25	N	Education	UK
26	HealthMedUK26	N	Health, medicine	UK
27	HealthNon-MedUK27	N	Health, non-medicine	UK
28	QAPG28	Y	Postgrad	UK
29	QAUG29	Y	Undergrad	UK
30	HealthMedINT30	N	Health, medicine	International
31	EducationUK31	N	Education	UK
32	EducationINT32	N	Education	International
33	QAPG33	Y	Postgrad	UK
34	HealthMedINT34	N	Health, medicine	International

Figure 3. Stakeholder groups

[Insert figure 3 here]

Main findings

We present findings which verify, refute and challenge our initial programme theory (themes 1-4), leading to the development of our modified programme theory. Contexts, mechanisms and outcomes are labelled as [C], [M], and [O] respectively. CMO configurations, resources and responses are identified and illustrated across themes.

Theme 1) Quality standards

Standards defined the level at which a provider needs to function to reach certain outcomes, e.g. meeting minimum standards. Key mechanisms triggered by co-construction of setting standards included compliance, clarification, flexibility and adherence. Undergraduate and postgraduate QAPs responded to the regulators standards by inclusion in their own policies. The standards also provided QAPs with leverage to push forward changes at institutions:

> '*I think, a very positive element [O], is that it [standards] has allowed UK medical schools [C] within the framework to differ in how they implement that framework…So I think as well as having the rigour of what must be done, it allows for a degree of flexibility [M].' (HealthMedINT1)*

> *'We reference them [the standards] in our…internal policies.' (QAUG19)*

However, the standards had unintended consequences as their presence sometimes created confusion, particularly in the postgraduate context due to lack of clarity. In some instances, organisations had their own standards to assess educational quality, resulting in confusion:

> *'I suspect [postgraduate organisation] ignores them [standards] because they've come up with their own quality framework.' (QAPG33)*

> *'Our biggest concerns [O] really are not so much the standards as the sort of processes that by which the GMC [C] will check that our curricula are... comply with those standards [M]. And I think on that, you know, looking back now I can see that there was a certain ambiguity [M] in how the GMC were going to approach this and I'm not sure that they ever resolved it as the standards were being developed [O].' (QARC3)*

Prescriptive, rigid and inflexible standards prevented providers from being adaptable to need and innovation. For example standards which focus on particular aspects (e.g. student diversity) detracted attention from other areas of need (e.g. widening participation). Conversely, less binding standards (e.g. not detailing specific teaching methods) triggered mechanisms of ambiguity, openness and flexibility creating too much variation in education across contexts and producing new risks to quality. Misalignment between different quality standards caused frustration. For example, a LEP was deemed to be clinically outstanding but was also found to be inadequate for educational quality by a different regulator. Local pressures were seen to inhibit postgraduate partners' abilities to follow standards, suggesting that in the 'real world', applicability of standards was sometimes questionable:

> *'A lot of LEPs take our students [C], but they [LEPs] can quite readily tell us [medical school] to take them away as well [O], if we're very strict with them about meeting certain standards and certain criteria [M].' (QAUG5)*

Theme 2) Sanctions and approvals

We identified that organisational culture affects approaches to sanctions, and so an 'acceptable sanction' was contingent on risk. However in different contexts, should supportive measures be inadequate then the most severe sanction of withdrawn approval should remain. The "ultimate sanction of power" (QAPG28) was regarded to fulfil its intended consequence i.e. to protect patient safety, but also had unintended consequences to reinforce the medical regulator's authority and subsequently motivate providers to address problems. There was a firm belief that a severe sanction should rarely need to be enforced if other QA components (e.g. self-assessment) are effective.

> *'It's [sanctions] a bit of a lightning rod situation, but I think it [closing medical school] should remain as the ultimate sanction [O]…If trust management realised for example that they wouldn't lose their trainees as a result of not providing a safe and effective*

training environment [M]… I think [it would] slip further down their list of priorities [O].' (QARC24)

The effectiveness of regulatory approvals in the undergraduate and postgraduate context varied. In undergraduate contexts, it was described as time-consuming examining both curriculum and staff capabilities. However, in the postgraduate context the process and need for QA was not as clear. Non-QAPs felt that it was important that mechanisms were in place to periodically review approvals. For instance:

'[The thoroughness] enabled them [regulator] to make a decision on our suitability to proceed.' (QAUG29)

'what if the trainee goes for one week, but it's only one week out of a one year placement, do they need to get that site approved?.' (QAPG12)

'We don't link approvals and quality very strongly [C]…we go to the GMC and we say, can we put some doctors here please? And the GMC go, yes. But there's an implication in doing that that because we're asking, we're going to quality manage that particular set of placements [M]. And we do, but not explicitly and not formally [O].' (QAPG16)

Theme 3) Collecting information: Visits, monitoring and self-assessment

Institutional visits were positioned as a key component as they triggered internal reviews and reflection, subsequently motivating organisations to improve quality. Working collaboratively engendered trust with open and honest dialogue, which was considered crucial in effecting change. Meaningful dialogue and collaboration were important and that was achieved through having high quality, "respected" (EducationINT32) trained visitor teams:

'I was prepared to be completely open and honest with the GMC…If [visits] are going to be effective, relationship building is actually more important than what you're doing collecting evidence.' (HealthMedUK21)

The QAF includes a range of monitoring data collection processes such as: data from the national training survey (NTS); monitoring including enhanced monitoring; self-assessment; and visits. The NTS surveys all doctors in training which facilitated increased transparency, accountability and risk-identification. The resources provided by the survey could lead to invaluable outcomes to examine training differences, make evidence-informed decisions and pinpoint training issues. However, multiple sources of data were sometimes regarded as conflictual, obscuring the overall picture of education quality.

'At the moment, they're [GMC] looking [at] trainee burnout [C]. So they're generating all this data at the moment and I don't think they're clear about what they're going to do with it [M], and my concern is they will just dump it on us for us to fix, and I don't think we can [O].' (QAPG10)

The component of requiring self-assessment triggered many different mechanisms in different contexts. For regulators, it generated reflective internally-led processes. The reasons for this were around connectivity between regulators and providers:

'[self-assessment] a really fundamental part of what we do, and we place a massive... emphasis on that.' (EducationUK17)

'this is the way [self-assessment] an institution connects itself with given standards" (EducationINT32)

'to work constructively with the provider…being the start of a peer review process' (HealthNon-MedINT11)

Whereas, for those who were being quality assured, the formality of written self-assessment inhibited open disclosure as it was laborious and seen more as an audit. Validity and reliability was also raised as perceptions of lack of feedback from the regulator also undermined self-assessment.

'I think you're more likely to hear genuine issues, genuine things that need to be fixed, if you speak to people informally and off the record' (QAPG6)

'It [self-assessment] forms part of it[assuring quality] and it's a very strong part of it, but I wouldn't necessarily use it [self-assessment] in isolation' (HealthMedUK23)

'*We are encouraging of institutions identifying challenges [C]. So if an institution is very open and honest [M], even into what might be quite a delicate area, saying this has been a challenge for us [C] and we're working away on it and we're doing the following things. Provided that their plan of action is a good one and that it's being conducted in a timely manner that would be reported on in a positive light [O].' (HealthNonMedUK9)*

Theme 4) Reporting; accountability, dissemination, good practice

A patient safety outcome response identified from external reporting was to build public confidence. Publicly available outputs fostered accountability to illustrate how providers are low-risk thus requiring less scrutiny. Insufficient reporting and feedback (in terms of timeliness, quantity and quality) fostered outcomes of devaluing time and effort, and subsequent disengagement. Risk context was also important to determine the effectiveness of intervention components. Informal partnerships were highlighted as critical to assuring quality.

'I think the transparency in publications are important because it involves or it makes things clear and open to all stakeholders.' (QAPG28)

'I think what having it public does, is it creates some pressure and accountability [M] on both the accreditor and the accrediting body [C] to focus on the outcomes and to show progress against conditions [O].' (HealthMedINT13)

'Working in partnership with regulators was instrumental, it has a significant effect on driving change in trusts' (HealthedUK21)

When feedback mechanisms triggered included collaboration and openness this fostered informal working partnerships leading to a raft of positive outcomes including awareness, sharing knowledge and quality improvement. Rapport over time helped provider's develop trust to report concerns. Here the positioning of regulator-provider context, moved from accountability quality-checker to collaborative problem-solver. Spreading good practice was contextually variable.

'The institution needs to take that genuine look at it, and spend the time genuinely evaluating and genuinely creating action plans…an institution that is good at critical self-reflection will tend to address problems before, or potential problems, before they become actual problems [O].' (EducationUK17)

'What works for one school may not work for other [O] …So, you don't want people to blindly be saying oh, let's do that, because that's going to be good practice here [M]….Because education in programs do vary [C]' (HealthNon-MedUK18)

Subsequently motivation reaction was low to implement changes based on other organisations' examples of good practice. Reports were deemed most helpful when they included action plans accessible to lay audiences.

DISCUSSION

Summary of findings

We undertook a realist evaluation to explore UK medical education QA. We found that intervention components support or undermine QA for different organisations, and at different times in undergraduate (for medical students) and postgraduate (for trainees) contexts. In the results we revealed that although interventions were often implemented uniformly in undergraduate and postgraduate contexts, the impact of these varied with some leading to positive and negative outcomes. We tested our initial programme theory (themes 1-4) to develop a modified programme theory.

Across the three-tier model we identified that the undergraduate and postgraduate context were influential. For example in the undergraduate context, the leverage brokered by the regulator was often associated with directive features enabling local changes. Whereas in the postgraduate context, this power was often lost and diffused across education layers. Predominantly in the postgraduate trainee context, interventions led to unintended consequences (e.g. organisation disengagement) if an intervention promoted adherence at the cost of autonomy, subsequently triggering a lack of motivation (theme 1). An underlying mechanism identified to ensure an inclusive approach to QA was partnership working (themes 3-4). In the undergraduate context, provider-regulator engagement was sometimes not present, typically when there was a lack of informal relationships. In theme 3, visits were identified as a component that could better foster partnerships; so long as they were conducted with integrity, meaning and purpose.

Collectively across themes 1-4, the QAF was commended for being comprehensive and enabling a broad understanding of an organisation's performance. Internally-led processes with organisations identifying and addressing their own challenges and deficiencies, when done well, promoted a sense of autonomy and accountability (theme 3). The main unintended QAF consequences fell broadly into two outcomes, those related to the overlap across the three tiers (themes 1-2) and those related to the regulatory burden associated with data-driven approaches with a lack of transparency on why and how data was used (themes 3-4). A blurring of roles and boundaries of multiple organisations between patient safety, medical education and training was identified.

The influence and power of the regulator was continually picked up across the components (themes 1-4) which triggered mechanisms including transparent reporting to promote quality improvement, effective communication, trust, and partnership working facilitating interactions between regulators, partners and providers. Proportionate reactions in the face of disclosing and identifying patient safety risks at an early stage were more likely to occur within a positive trusting regulator-provider context underpinned by openness. Likewise, an organisation that self-assessed critically was reported to give regulators confidence in the institution.

Modified programme theory

The findings informed a modified programme theory to explain the underlying processes for the intended and unintended consequences of how the GMC quality assure education in various contexts (Figure 4). The modified programme theory conceptualises CMO configurations presented so that components are understood in the ways in which they may lead to positive or negative outcomes. The programme theory broadly conceptualises the differing QAPs contexts as having associative or dissociate contextual features.

Associative contextual features are exhibited where the QAP demonstrates adherence to the regulator's QAF, either in full or to particular components, triggering positive mechanisms and outcomes. Dissociative contextual features are where there is organisational resistance to imposed external QA triggering negative mechanisms and outcomes. Each of the QAF components can therefore be enacted differently within these different contexts. Across the three-tier model (quality assurance, quality management and quality control), the theory demonstrates that QAF components are enacted differently depending on whether the context has associative or dissociative features, with implications for the mechanisms triggered leading to positive (e.g. effecting change, contextual application of standards, partnerships) and/or negative outcomes (limited compliance, resistance, overlap).

Figure 4. Modified programme theory to explain the underlying process for the intended and unintended consequences of quality assurance components

[Insert figure 4 here]

Relevance of findings and implications

The findings reinforce the quality assurance literature highlighting *trust* in fostering effective working relationships to enhance feedback.[1 5] We extend this further, and identified that early communication of emerging risks supports quality assurance and enhancement approaches through informal networks. Visits aid communication and build relationships, yet if lost may distance the regulator and undermine opportunities for partnership working. Informal communication provided a safe environment for providers to discuss concerns with the regulator, opposed to formal monitoring acting as a barrier.

Expanding the literature, we demonstrate that context must be considered in order for quality assurance to protect patients.[23] Risk is context dependent, and was perceived to be tangibly different across undergraduate and postgraduate contexts. Undergraduate medical settings were perceived as low risk and imply opportunities for greater tailoring and focus. The overlaps between quality assurance, quality management and quality control were apparent especially within the postgraduate setting with duplication and confusion of responsibilities. These findings align with a recent systematic review identifying features of failing healthcare organisations including conflicting missions, fragmented accountability and lack of collaboration.[1]

Collectively, this supports the need to clarify structural quality processes and how organisations are intended to function collaboratively. In the analysis, risk-based visiting positioned the regulator as quality assurer rather than quality enhancer. Equally, effective assurance is often associated with suppressing innovation.[24] Moreover, the role of self-assessment[24] posed a number of challenges in relation to purpose and autonomy. While institutional self-assessments can positively influence reactions to drive quality improvement, there are issues with validity, reliability and internal quality review.[19 24-27]

The power of the regulator impacted on the effectiveness of intervention components in multiple ways. The regulatory-burden associated with monitoring activities was considerable and disengagement ensued. Lack of feedback from the regulator was an important aetiological mechanism precipitating the situation. Similarly, negative consequences of approvals including cost, low staff morale, threats to organisation reputation, and the suppression of innovation through adhering to standards has been identified.[28 29] Without regulators addressing varying risk contexts, the proportionality of QA is imbalanced, leading to negative outcomes with regulators unable to effectively assure quality. Therefore, collectively considering a hybrid model of cyclical plus risk-based visiting may help to build provider relationships and drive improvement whilst ensuring minimum standards. Collective assurance and relationships should be encouraged so that regulators and providers can tackle issues conjointly. Flexibility in utilising other datasets within any collaborative work is a necessity and a clear stance on organisational remit and particularly boundaries, is anticipated to be a key mechanism in effective joint QA.

Strengths, limitations and future directions

To our knowledge this is the first robust study on education QA within the healthcare context, synthesising data from stakeholders. The study fills a gap as QA remains expensive, yet its functionality is largely unexplored. The study was conducted by an experienced multidisciplinary research team applying an innovative realist approach, underpinned by a sound team-based analysis. A somewhat surprising omission from our findings is a lack of attention to the mechanism of leadership.[1 23] The sample focussed on processes rather than delivery perhaps contributing to such omission. Moving forward, there is a need to conduct an economic review and consult with stakeholders into what data could be shared (e.g. National Training Survey, Care Quality Commission data) to understand links to intervention components. The findings have influenced the GMC's approach to QA which impacts on all medical students and doctors in training.[30]

Conclusions

This study used a realist methodology to reveal the intended and unintended consequences of components used by the GMC to quality assure medical education, and elucidated the mechanisms by which both are brought about. While uniform approaches are often in place, interventions need to be contextually tailored. Ongoing partnership working can enhance open disclosure to drive up education quality. This research has provided a modified programme theory to explicate how education providers and regulators can work more effectively together to uphold quality, and ultimately protect public safety.

Acknowledgements

We would like to thank the participants for sharing their insights, the GMC for facilitating recruitment, and Dr Graham Easton and UCL Design team for helping to produce the informational video.

Author contributions

AG, PC, LM and MP conceived the study design. PC and LM designed study materials and gained ethical approval. All authors collected data, contributed to data analysis. PC wrote the first draft of the manuscript with input from AG, MP, LK and LM. All authors approve the final manuscript.

Data sharing

We do not have ethical permission to share raw data.

Funding and conflict of interest

This study was commissioned by the GMC, project ID: GMC 822. AG was the principal investigator.

Ethics

This study was registered with UCL's data protection office on 06/06/18 and approved by UCL ethics on 15/06/18, project ID: 6281/003. All participants volunteered to take part and provided written consent. Data were anonymised with respect to individuals and institutions.

References

1. Vaughn VM, Saint S, Krein SL, et al. Characteristics of healthcare organisations struggling to improve quality: results from a systematic review of qualitative studies. *BMJ Qual Saf 2019*, *28*(1), pp.74-84.

2. General Medical Council. Promoting excellence 2019. Accessed on 16.07.19 from: https://www.gmc-uk.org/education/standards-guidance-and-curricula/standards-and-outcomes/promoting-excellence

3. Vroeijenstijn, A.I. Quality assurance in medical education. *Academic medicine: journal of the Association of American Medical Colleges* 1995, *70*(7 Suppl), pp.S59-67.

4. General Medical Council (2019). How we assure quality. Accessed on 10/04/2019 from: https://www.gmc-uk.org/education/how-we-quality-assure

5. Jones K & Tymms P. Ofsted's role in promoting school improvement: the mechanisms of the school inspection system in England, *Oxford Review of Education 2014*, 40(3), 315-330, DOI: 10.1080/03054985.2014.911726

6. Westerheijden DF, Stensaker B, Rosa MJ (eds). *Quality assurance in higher education: Trends in regulation, translation and transformation* (Vol. 20) 2007. Springer Science & Business Media.

7. Haji F, Morin MP, Parker K. Rethinking programme evaluation in health professions education: beyond 'did it work?'. *Medical education 2013*, *47*(4), pp.342-351.

8. Øvretveit J, Gustafson D. Evaluation of quality improvement programmes. *BMJ Qual Saf* 2002;11:270-275.

9. Pawson R, Greenhalgh T, Harvey G, et al. Realist review – a new method of systematic review designed for complex policy interventions, *Journal of Health Services Research & Policy 2005* 10(1): 21–34

10. Pawson R. *The Science of Evaluation: A Realist Manifesto*, SAGE Publications Ltd, London, 2013.

11. Jones AC, Shipman SA, Ogrinc G. Key characteristics of successful quality improvement curricula in physician education: a realist review BMJ Qual Saf 2015;24:77-88

12. Wong G, Westhorp G, Manzano A, et al. RAMESES II reporting standards for realist evaluations. *BMC medicine 2016*, *14*(1), p.96.

13. Griffin A, Crampton P, Mehdizadeh L, et al. Understanding stakeholder perspectives on the GMC's quality assurance of medical education and training. Final Report, 2018. Accessed 25.07.19 from: https://www.gmc-uk.org/-/media/documents/understanding-stakeholder-perspectives-on-the-gmcs-quality-assurance-of-medical-education-and-traini.pdf?la=en&hash=14529235D9E6C99AF9B5EDA82F6848CB34D33396

14. Jagosh J, Bush PL, Salsberg J et al. A realist evaluation of community-based participatory research: partnership synergy, trust building and related ripple effects. *BMC public health 2015*, *15*(1), p.725.

15. Manzano A. The craft of interviewing in realist evaluation. Evaluation. 2016 Jul;22(3):342-60.

16. Keren G, De Bruin WB. On the assessment of decision quality: Considerations regarding utility, conflict and accountability. Thinking: Psychological perspectives on reasoning, judgment and decision making. 2003 Sep 16;347.

17. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. *The qualitative researcher's companion* 2002; 573, 305-329.

18. NVivo qualitative data analysis software; QSR International Pty Ltd. Version 11, 2015.

19. Griffin A, McKeown A, Viney R, et al. Revalidation and quality assurance: the application of the MUSIQ framework in independent verification visits to healthcare organisations. *BMJ Open* 2017;7:e014121. doi:10.1136/bmjopen-2016-014121

20. Griffin A, Knight L, McKeown A, et al. A postgraduate curriculum for integrated care: a qualitative exploration of trainee paediatricians and general practitioners' experiences. *BMC Med Educ* 2019; *19*(1), p.8.

21. Rees CE, Crampton PES, Kent F, et al. Understanding students' and clinicians' experiences of informal interprofessional workplace learning: an Australian qualitative study. BMJ open 2018.1;8(4):e021238.

22. Ajjawi R, Crampton PES, Rees CE. What really matters for successful research environments? A realist synthesis. *Medical education 2018*, *52*(9), pp.936-950.

23. Kaplan HC, Provost LP, Froehle CM, & Margolis PA. The Model for Understanding Success in Quality (MUSIQ): building a theory of context in healthcare quality improvement. *BMJ Qual Saf 2012*, *21*(1), pp.13-20.

24. Lang DW. Self-regulation with rules: Lessons learned from a new quality assurance process for Ontario, *Quality Assurance in Education 2015*, 23:3, 216-232, https://doi.org/10.1108/QAE-09-2014-0046

25. Tackett S, Grant J & Mmari K. Designing an evaluation framework for WFME basic standards for medical education, *Medical Teacher 2016*, 38(3), 291-296, DOI:10.3109/0142159X.2015.1031737

26. Jones KL, Tymms P, Kemethofer D, et al. The unintended consequences of school inspection: the prevalence of inspection side-effects in Austria, the Czech Republic, England, Ireland, the Netherlands, Sweden, and Switzerland, *Oxford Review of Education 2017*, 43(6), 805-822, DOI: 10.1080/03054985.2017.1352499

27. Forrest C. Inspection and improvement in three further education colleges, *Research in Post-Compulsory Education 2015*, 20(3), 296-314, DOI: 10.1080/13596748.2015.1063273

28. de Paor C. The contribution of professional accreditation to quality assurance in higher education, *Quality in Higher Education 2016*, 22(3), 228-241, DOI: 10.1080/13538322.2016.1263925

29. Blouin D, Tekian A, Kamin C, et al. The impact of accreditation on medical schools' processes, *Medical Education* 2018, 52(2), 182-191

30. General Medical Council. Council Papers February 2019. Accessed on 17.07.19 from: https://www.gmc-uk.org/-/media/documents/council---agenda-and-papers-for-meeting-on-27-february-2019_pdf-78924314.pdf

Figure 1: The intervention: quality assurance framework (QAF)

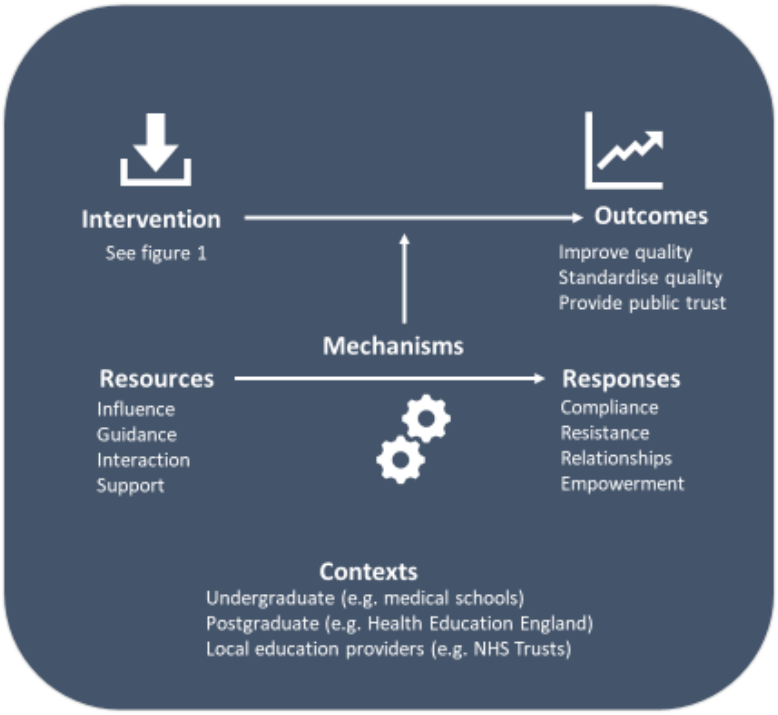


Figure 2. Initial programme theory

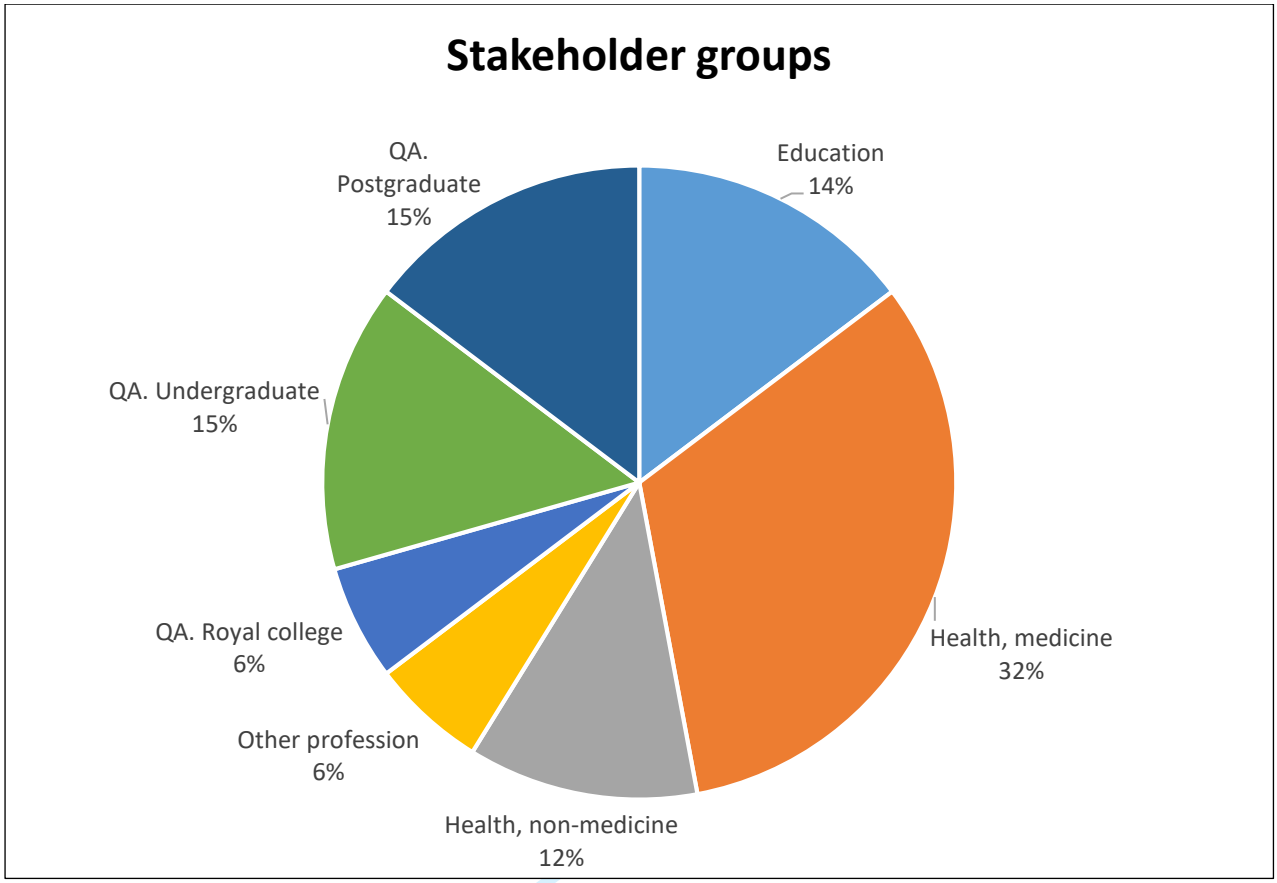
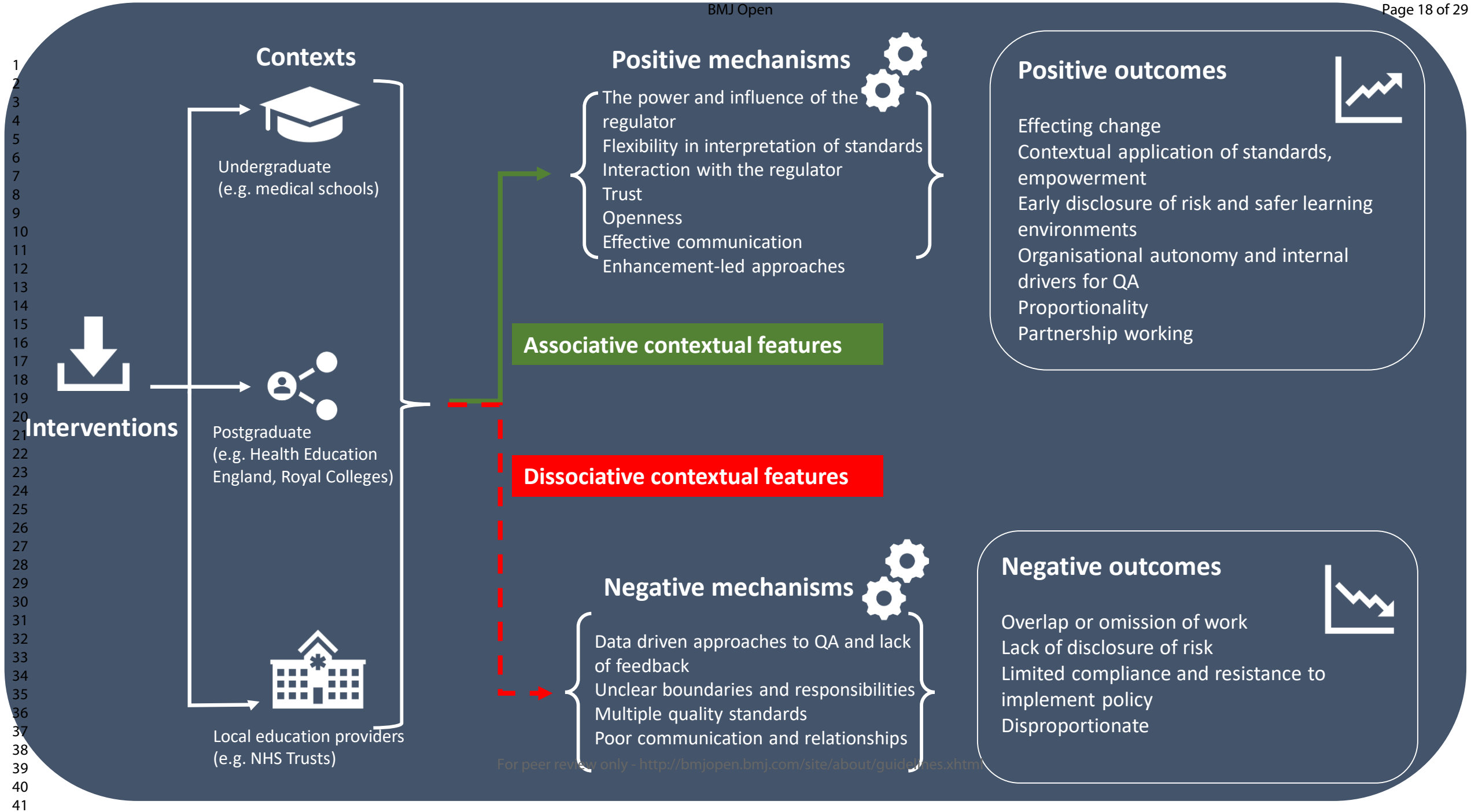


Figure 3. Stakeholder groups



1. Quality Assurance partners interview schedule

Structure

Introduce and duration about 45 mins

Process

This interview will be recorded and analysed, looking for common themes that arise across the interviews.

Consent

A reminder that we will not personally name anyone who takes part in the study. Do you have any questions before we start the tape? Thank you for signing the consent form [or take verbal consent if required].

Introductory/ Background questions

- How long have you worked in your quality assurance role?
- Who else in your organisation works in a QA role?

General: GMC quality assurance framework overall

- Are there any aspects of GMC's quality assurance framework that you think are particularly effective, i.e. give you reassurance in their processes?
- Are there any aspects of the framework that you think are less effective or problematic in some way, i.e. do not assure you?

Focused: Specific aspects of GMC's quality assurance framework

I would now like to ask some questions about different components of the QA framework and be keen for you to share your experiences where relevant.

Standards

- Are the standards the right ones? Prompt any missing?
- Are the standards helpful or unhelpful in anyway?
- Has using the standards had any impact on your organisation?

Approvals

- What would be the advantages of making the GMC's approvals time limited?
- What would be the disadvantages of making the date GMC's approvals time limited?
- Do you think the GMC's approvals process is effective?

Monitoring:

- Is the GMC monitoring the right evidence to assess your organisations performance?

- What sources of evidence do you think give the GMC the best insight into your organisation?
- What other areas could they/should they monitor?
- Does monitoring have any impact on your organisation? Prompt: Positive/negative
- Turning to enhanced monitoring, some people would say that the GMC are overstepping their remit when they require postgraduate organisations to report training programs and local education providers to them, what are your thoughts?

Sharing evidence:

- How effective is the GMC at sharing evidence with you?
- Is there evidence that could be shared more effectively and how would that benefit your organisation?
- Is there any evidence that you feel should not be shared, in particular with other healthcare regulators?

Self-assessment:

The GMC requires annual self-assessment from the medical Royal colleges and medical schools but not the postgraduate organisations.

- Do you think self-assessment is a helpful process?
- Some hold the view that organisational self-assessment is not a reliable process, what do you think?
- Has the process of self-assessment resulted in any organisational change?

Visits:

- What purpose do you think visits to organisations have?
    - Prompt: What makes a visit effective?
    - Prompt: What are the important areas that visits should include?
- Most regulators are moving away from cyclical or scheduled visiting, towards entirely risk-based systems, however many GMC stakeholders believe the cyclical visits have many benefits and should be retained. What do you think?
- What would happen if the GMC did no visiting?

Reporting:

The GMC currently publishes long-form visit reports on its website, as well as information about enhanced monitoring cases, and data tools such as the NTS reporting tool and the progression reports.

- What do you think of the current QA reporting?
- Are there any negative consequences of reporting data on the website?
- Does your organisation use the reports in anyway?

Good practice:

The GMC aim to identify good practice across medical schools and postgraduate bodies and then publish this on their website.

- Is this useful to your organisation?
    - Prompt: positive aspects v negative

- Have you adopted any areas of good practice yourself?
- Some people would say more resources should go into quality enhancement rather than accountability. What are your views?

Fairness

- How can the GMC quality assure fairness in medical education and training?

Sanctions:

Sanctions mean withdrawing trainee doctors from the NHS or closing down medical schools which has a critical impact on healthcare.

- In the case of an underperforming training organisation that is currently failing to meet required standards what might be a proportionate sanction from the GMC that is not as extreme as withdrawing approval?
    - Prompts: The GMC visiting, publicly available rating scales, time bound approvals

Collective assurance

The GMC has committed to working more closely with other regulators to find efficiencies and reduce the regulatory burden on the service.

- What would be the advantages for your organisation in this approach?
- Would there be any disadvantages?
- Would sharing data enable the GMC to identify risk better?
- How practical would it be for your organisation to undertake joint visiting?
- Do you think the GMC's approach to QA is proportionate to the risks involved in medical education and training?
- Do you have any suggestions for improving the GMC's quality assurance processes?

Thank you for your time. Is there anything you would like to add that we haven't discussed?

Thank you

2. Quality Assurance non-partners interview schedule

Structure

Introduce and duration about 45 mins

Process

This interview will be recorded and analysed, looking for common themes that arise across the interviews.

Consent

A reminder that we will not personally name anyone who takes part in the study. Do you have any questions before we start the tape? Thank you for signing the consent form [or take verbal consent if required].

Introductory/ Background questions

- Can you briefly explain the context in which your organisation is involved in QA
- What is your specific role?
- How long have you worked in your quality assurance role?
- Who else in your organisation works in a QA role?

General: GMC quality assurance framework overall

- Are there any aspects of GMC's quality assurance framework that you think are particularly effective, i.e. give you reassurance in their processes?
- Are there any aspects of the framework that you think are less effective or problematic in some way, i.e. do not assure you?

Focused: Specific aspects of GMC's quality assurance framework

I would now like to ask some questions about different components of the QA framework and be keen for you to share your experiences where relevant.

Standards

- Are the standards the right ones?
    - Prompt: Any missing?
- Are the standards helpful or unhelpful in anyway?

Approvals

- Do you think the GMC's approvals process is effective?
- What would be the advantages of making the GMC's approvals time limited?
- What would be the disadvantages of making the date GMC's approvals time limited?

Monitoring

- Is the GMC monitoring the right evidence to assess organisational performance?
    - Prompt: What other areas could they/should they monitor?
- Do you think monitoring has any impact on organisation performance?
    - Prompt: Positive/negative
- Turning to enhanced monitoring, some people would say that the GMC are overstepping their remit when they require postgraduate organisations to report training programs and local education providers to them, what are your thoughts?
- How does your organisation use monitoring?
- Do you have a model for triangulating predicting risk?

Sharing evidence

- How could the GMC improve sharing its evidence?
    - Prompt: Between regulator to regulator; between regulators to QA partners?
- Is there other evidence that could be shared?
    - Prompt: Is there any evidence that you feel should not be shared?

Self-assessment

The GMC requires annual self-assessment from the medical Royal colleges and medical schools but not the postgraduate organisations.

- Do you think self-assessment is a helpful process?
- Some hold the view that self-assessment is not a reliable process, what do you think?
- What is your organisations approach to self-assessment?
- Has the process of self-assessment resulted in any organisational change?

Visits

- What purpose do you think visits to organisations have?
    - Prompt: What makes a visit effective?
    - Prompt: What impact do you think organisational visits have?
- Most regulators are moving away from cyclical or scheduled visiting, towards entirely risk-based systems, however many GMC stakeholders believe the cyclical visits have many benefits and should be retained. What do you think?
- How can visits give greater assurance of quality?
- What would happen if the GMC did no visiting?

Reporting

The GMC currently publishes long-form visit reports on its website, as well as information about enhanced monitoring cases, and data tools such as the NTS reporting tool and the progression reports.

- What do you think of the GMC approach to reporting?
- How does your organisation report on performance?
    - Prompts: strengths and weaknesses?

Good practice

The GMC aim to identify good practice across medical schools and postgraduate bodies and then publish this on their website.

- What do you think of the GMCs approach to sharing best practice?
- Some people would say more resources should go into quality enhancement rather than accountability. What are your views?
- What is your organisations approach to this?

Fairness

- How can the GMC quality assure fairness in medical education and training?

Sanctions

In the GMC's context, sanctions mean withdrawing trainee doctors from the NHS or closing down medical schools which has a critical impact on healthcare.

- Does your organisation have any advice or experience of imposing meaningful sanctions that would not be considered as extreme as the GMC's approach?

Collective assurance

The GMC has committed to working more closely with other regulators to find efficiencies and reduce the regulatory burden on the service.

- Is your organisation involved in joint visits? If so, what would be the advantages/ disadvantages for your organisation in this approach?
- Do you sharing data with other organisations?
- How practical would it be for your organisation to undertake joint visiting?
- Do you think the GMC's approach to QA is proportionate to the risks involved in medical education and training?
- Do you have any suggestions for improving the GMC's quality assurance processes?

Thank you for your time. Is there anything you would like to add that we haven't discussed?

Thank you

RAMESES 2 reporting checklist

Source: Wong, G., Westhorp, G., Manzano, A., Greenhalgh, J., Jagosh, J. and Greenhalgh, T., 2016. RAMESES II reporting standards for realist evaluations. *BMC medicine*, *14*(1), p.96.

TITLE			**Reported in document**	**Section, Page(s) in document**
1		In the title, identify the document as a realist evaluation	Y	Title, pg 1
SUMMARY OR ABSTRACT				
2		The abstract or summary should include brief details on: the policy, programme or initiative under evaluation; programme setting; purpose of the evaluation; evaluation question(s) and/or objective(s); evaluation strategy; data collection, documentation and analysis methods; key findings and conclusions Where journals require it and the nature of the study is appropriate, brief details of respondents to the evaluation and recruitment and sampling processes may also be included Sufficient detail should be provided to identify that a realist approach was used and that realist programme theory was developed and/or refined	Y	Abstract, pg 2
INTRODUCTION				
3	Rationale for evaluation	Explain the purpose of the evaluation and the implications	Y	Introduction pg. 3-4

TITLE			Reported in document	Section, Page(s) in document
		for its focus and design		
4	Programme theory	Describe the initial programme theory (or theories) that underpin the programme, policy or initiative	Y	Methodology pg 4
5	Evaluation questions, objectives and focus	State the evaluation question(s) and specify the objectives for the evaluation. Describe whether and how the programme theory was used to define the scope and focus of the evaluation	Y	Introduction, study aim, pg 3-5
6	Ethical approval	State whether the realist evaluation required and has gained ethical approval from the relevant authorities, providing details as appropriate.	Y	Ethical approval, pg 5 All participants volunteered to take part and provided written consent.
METHODS				
7	Rationale for using realist evaluation	Explain why a realist evaluation approach was chosen and (if relevant) adapted	Y	Methods, conceptual framework, pg 4-5
8	Environment surrounding the evaluation	Describe the environment in which the evaluation took place	Y	Methods, conceptual framework, data collection, pg 4-6
9	Describe the programme policy, initiative or product evaluated	Provide relevant details on the programme, policy or initiative evaluated	Y	Introduction, pg 3-4; Methods, conceptual framework, Pg 5-6
10	Describe and	A description and justification	Y	Introduction, pg

TITLE			Reported in document	Section, Page(s) in document
	justify the evaluation design	of the evaluation design (i.e. the account of what was planned, done and why) should be included, at least in summary form or as an appendix, in the document which presents the main findings. If this is not done, the omission should be justified and a reference or link to the evaluation design given. It may also be useful to publish or make freely available (e.g. online on a website) any original evaluation design document or protocol, where they exist		4; Methods, conceptual framework, Pg 5-6
11	Data collection methods	Describe and justify the data collection methods – which ones were used, why and how they fed into developing, supporting, refuting or refining programme theory Provide details of the steps taken to enhance the trustworthiness of data collection and documentation	Y	Methods, data collection, pg 4-5
12	Recruitment process and sampling strategy	Describe how respondents to the evaluation were recruited or engaged and how the sample contributed to the development, support, refutation or refinement of programme theory	Y	Methods, sampling, recruitment, pg 4-5
13	Data analysis	Describe in detail how data	Y	Methods,

TITLE			Reported in document	Section, Page(s) in document
		were analysed. This section should include information on the constructs that were identified, the process of analysis, how the programme theory was further developed, supported, refuted and refined, and (where relevant) how analysis changed as the evaluation unfolded		analysis, reflexivity, pg 5
RESULTS				
14	Details of participants	Report (if applicable) who took part in the evaluation, the details of the data they provided and how the data was used to develop, support, refute or refine programme theory	Y	Results, participant details, pg 5
15	Main findings	Present the key findings, linking them to contexts, mechanisms and outcome configurations. Show how they were used to further develop, test or refine the programme theory	Y	Results, main findings, pg 5-9
DISCUSSION				
16	Summary of findings	Summarise the main findings with attention to the evaluation questions, purpose of the evaluation, programme theory and intended audience	Y	Discussion, summary of key findings, pg 9-11
17	Strengths, limitations and future directions	Discuss both the strengths of the evaluation and its limitations. These should include (but need not be limited	Y	Discussion, Strengths, limitations and future directions, pg

TITLE			Reported in document	Section, Page(s) in document
		to): (1) consideration of all the steps in the evaluation processes; and (2) comment on the adequacy, trustworthiness and value of the explanatory insights which emerged The particular implications arising from the realist nature of the findings should be reflected in these discussions		11-12
18	Comparison with existing literature	Where appropriate, compare and contrast the evaluation's findings with the existing literature on similar programmes, policies or initiatives	Y	Discussion, Comparison with existing literature and implications, pg 10-12
19	Conclusion and recommendations	List the main conclusions that are justified by the analyses of the data. If appropriate, offer recommendations consistent with a realist approach	Y	Discussion, implications, conclusions, pg 10-12
20	Funding and conflict of interest	State the funding source (if any) for the evaluation, the role played by the funder (if any) and any conflicts of interests of the evaluators	Y	Funding and competing interests, pg 12 This study was externally commissioned by the GMC