Centre for Education Systems

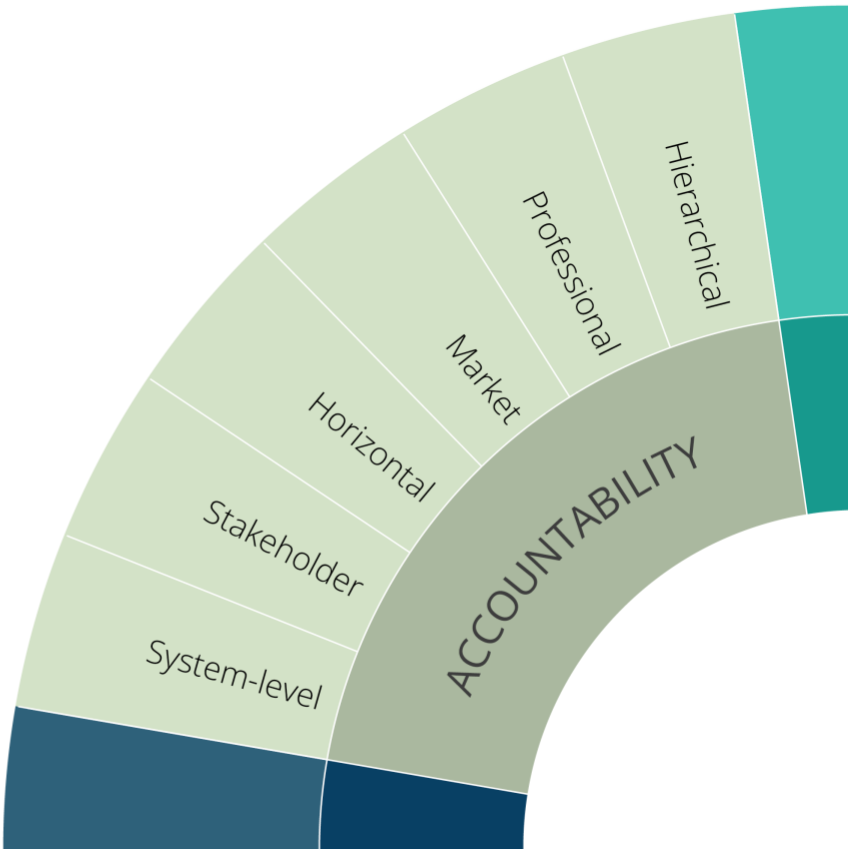
INTERNATIONAL COMPARATIVE REVIEW

ACCOUNTABILITY POLICY

Report

Authors: Loic Menzies, Kusha Anand, Simon Burgess, Lucy Crehan, Melanie Ehren, Jessica Ko Sum Yue, Bevil Luck, Alison O'Mara-Eves, Antonia Simon, Carol Vigurs, Patrick Wall

November 2025 v1.0



About the Centre for Education Systems (CES)

The Centre for Education Systems improves education for all, by helping governments, advisors and campaigners learn from education systems around the world. We describe different policy approaches, their contexts, and their outcomes.

See www.CES.partners

Acknowledgements

The EPPI Centre

This report is based on a comparative review undertaken by the EPPI Centre – the Evidence for Policy and Practice Information Centre, based at University College London (UCL), within the Social Research Institute. EPPI conducts and develops systematic reviews to inform public policy in areas like health and education. Its mission is to promote rigorous evidence for a fairer world through methods for evidence synthesis, user participation in research, and a focus on research impact. To find out more about the work of the EPPI Centre, please visit our website eppi.ioe.ac.uk.

Nuffield Foundation

The Nuffield Foundation is an independent charitable trust with a mission to advance social well-being. It funds research that informs social policy, primarily in Education, Welfare and Justice. The Nuffield Foundation is the founder and cofunder of the Nuffield Council on Bioethics, the Ada Lovelace Institute and the Nuffield Family Justice Observatory. The Foundation has funded this project, but the views expressed are those of the authors and not necessarily the Foundation.

www.nuffieldfoundation.org

The Education Policy Institute (EPI)

The Nuffield Foundation grant was administered via EPI and EPI has provided support to CES throughout the programme.

www.epi.org.uk

Contents

Key takeaways **4**

Summary **7**

Summary References 13

Part 1: Introducing accountability **15**

Part 2: The characteristics of accountability systems **20**

RQ1: Purpose, values and trends 20

RQ2: The structure of accountability systems 24

Part 2 References 61

Part 3: The development and evaluation of accountability policy **64**

RQ3: Policy design 64

RQ4: Policy review and evaluation 68

Part 3 References 72

Part 4: The contexts for, and effects of accountability policy **73**

RQ5: Accountability contexts 73

RQ6: The effects of accountability systems 76

Part 4 References 91

Part 5: Overarching Findings **99**

Typical approaches 99

Part 5 References 119

CMO Configurations 125

Appendix 1: Methods **129**

Work Package 1: Led by Alison O'Mara-Eves and Antonia Simon 132

Work package: Led by Simon Burgess 133

Work package 3: Led by Melanie Ehren 133

Work Package 4 (this report): Led by Loic Menzies 134

Appendix 2: Concept note **135**

Complete Reference List 143

Key takeaways

1. There are seven main ways accountability systems attempt to generate improvement.

I. **By deterring malpractice**: This requires clear expectations with a credible possibility of sanctioning. It is primarily suited to ensuring compliance with minimum standards.

II. **By prompting support that builds expertise and capacity:** This can involve directing additional resources towards struggling schools and usually requires school-improvement infrastructure (e.g. networks of school advisors). Guidance needs to be of high quality and needs to be given in the right conditions, to professionals who are capable of making the necessary changes.

III. **By encouraging professionals to reflect on their practice and to generate solutions:** This can involve frameworks for self-evaluation and professional collaboration. This approach is likely to depend on a high-trust environment with skilled professionals.

IV. **Driving pupils towards better performing schools:** This requires parents to make informed choices about schools. It therefore depends on the availability of good alternatives and of accurate, accessible information about things that are commonly judged valuable. There is a high risk of inequitable outcomes and that struggling schools may find it even harder to improve.

V. **Prompting changes of management:** This can involve contractual arrangements with different school operators. It depends on the availability of effective managers, and conditions in which leaders can deliver changes.

VI. **By directing additional effort and attention towards government priorities:** This depends on incentives and pressures aligning with policy priorities. Professionals need capacity to direct additional efforts towards priorities without this causing problematic levels of stress and hyper compliance. Conformity with government priorities may mean other goals are neglected and innovation is inhibited.

VII. **Creating system alignment and coherence:** This requires shared priorities and professional learning across the system. Professionals need opportunities to interact and co-ordinate, as well as capacity to engage.

2. Accountability systems can focus on actors at different levels – including teachers, school leaders, governing bodies, or school operators. This has implications for who feels pressure in the system. Different elements of the accountability system can focus on each actor, for example peer networks might be used to hold teachers to account, while inspection might concentrate on governing bodies or school operators.

3. Different approaches to accountability fit with different governance arrangements: Some systems take a centralised approach. Some delegate to local governments. Others hand authority to intermediaries such as regulators. Preferences often reflect contextual factors, for example centralisation may be easier in small jurisdictions, whereas less state-centric approaches may suit jurisdictions in which religious schools' independence needs to be protected for historic reasons, or where the state wants to take a less active role in running public services.

4. Most jurisdictions want their schools to be responsive to the public, and accountability systems play a role in this.[1] The mechanisms through which they do this can be:

- **Electoral**: through the direct election of school boards, or through local government.
- **Participatory**: through joint self-evaluation and planning, or representation on governing bodies and councils.
- **Market-based**: through school choice.

5. Performance-based accountability – in which schools are held accountable for the outcomes pupils achieve – can lead to improvement, but it also has other predictable effects which may be unwanted. Well-documented effects include 'curriculum narrowing' and 'teaching to the test; pressure on teachers which can drive workforce turnover; and pupil exclusion and segregation. However, some of these effects can be intended – for example policy makers may want schools to prioritise certain subjects or content, or for less effective teachers to move on.

6. The specific design of an accountability system sets the incentives and shapes behaviour, but contextual factors can also shape responses. Inspection frameworks, and performance measures wire priorities into the system but schools respond in more or less learning-orientated or compliant ways depending on their time, expertise, and capacity to respond**.**[2]

7. Accountability systems need to build a coherent web of approaches that:

- **Balance two types of goals:**
    - ⊄ **A focus on minimum standards and the 'basics':** typically through compliance checks, measures that focus on core subjects or skills, and more frequent inspections for schools judged 'at risk'.
    - ⊄ **Expansive attempts to promote 'improvement for all':** involving wider educational priorities and continuous-improvement strategies.
- **Take into account combined effects including:**

[1] Kogan (1986)
[2] Ehren (2025)

- **Cumulative effects:** even if one element of the system has minimal consequences, these might prove significant once combined with other elements.
- **Interaction effects:** How teachers respond to one element of the system can be conditioned by other elements. For example the way teachers interact within a peer network is shaped by how their school will be judged in an inspection.

These findings raise seven key questions for policy makers

1. What is the underlying theory of improvement?
2. On whom should accountability pressures fall, and how much pressure is acceptable or desired?
3. What is the preferred, or existing approach to system governance (e.g. the degree of centralisation), and how will the accountability system fit with this?
4. Should responsiveness be built into accountability through electoral, market-based, or participatory means?
5. What guardrails are needed to protect against unintended but predictable consequences like pupil exclusion?
6. What should schools and professionals be incentivised to focus on, and how much capacity do they have to respond in intended ways?
7. How coherent is the accountability web? For example, are different elements of the system pulling in contradictory directions and what is the right balance between tackling underperformance versus pursuing more expansive goals?

Summary

The accountability conundrum

Accountability policy is highly contested. It has the potential to raise the performance of education systems but there is widespread concern that some accountability structures have a malignant effect.

This programme of work

This report documents fourteen jurisdictions' approaches to accountability and highlights some of the key choices to be made when developing policy in this area. Methods are described in Appendix 1.

Approaches to accountability

We identify seven main approaches to accountability. Jurisdictions combine elements of different approaches but tend to lean more heavily on some approaches than others.

Approach 1: Chain of command

This approach involves actors being accountable through a step-by-step chain – for example from Head, to school board, to regional authority, to national government. This normally involves managerial relations, but contractual and legal arrangements that focus on compliance with minimum standards (such as employing qualified teachers or safeguarding arrangements) can also play a role.

Approach 2: Vertical teacher accountability

This approach involves teachers being directly accountable to bodies outside of the school, skipping the 'chain of command'. External teacher accountability can focus on compliance, or be development focused.

Approach 3: Polycentric accountability

This approach involves multiple intermediary institutions, and legal and contractual relations typically play an important role. Actors can include school proprietors or operators, as well as independent or quasi-independent accountability bodies.

Approach 4: Market accountability

This approach involves the public holding school leaders or boards to account through market or choice-based approaches. Like polycentric accountability, this approach is intended to shift authority away from the state, but in market accountability, responsibility

for judgements shifts to parents and carers. Consequences are then expected to follow through the 'invisible hand' of aggregated choices.

Approach 5: Participatory accountability

This approach involves stakeholders (including parents, staff and the local community) directly holding school leaders or other managing authorities to account.

Approach 6: Horizontal professional accountability

This approach focuses on professional practice. Like in external teacher accountability, the actor is the teacher rather than the institution, but in this case peers hold each other to account, either internally or in networks and clusters.

Approach 7: Collective system accountability

This approach does not fall under narrower definitions of accountability. It involves the whole system being held to account collectively, in order to maximise performance at a jurisdiction level. It therefore encompasses teachers, schools, municipalities and government – and potentially even the public.

Purposes

Accountability systems differ more in form than purpose. Most of the reviewed accountability systems are seeking to:

What most accountability systems are seeking to do

Drive educational improvement or excellence;

Promote fairness, equity or inclusion;

Ensure compliance with government policy and regulation (particularly with regard to safeguarding);

Provide transparency and facilitate stakeholder or public engagement;

Promote well-being (of pupils and/or teachers);

Provide data or evidence to support decision making.

Four shifts in approach recur across numerous jurisdictions and are linked to changing policy priorities.

Four shifts in approach/changing policy priorities

1. A broadening of measures in response to a perceived 'narrow' focus on academic outcomes.
2. More 'risk-based', or 'differentiated' approaches to inspection that concentrate more on raising schools above a threshold, or which reduce the burden (and cost) of accountability.
3. Increased emphasis on diversity, equity and inclusion.
4. Prioritisation of workforce sustainability and professional learning.

The characteristics of accountability policy

Participants

Accountability systems are made up of 'actors' who are held to account and 'forums' who hold them to account. Many participants in the education system combine both roles.

Teachers are normally held to account by their managers within school, but in some systems they are held directly to account, either by the ministry, its inspectors or a professional body.

Other systems have structures or processes in place to help teachers hold each other to account, however these are often localised and unsystematic.

There are three main groups of jurisdictions, based on who participates in the accountability system:

1. **Highly centralised** systems with few, if any, non-governmental bodies where accountability is to the ministry or its appointed representatives and there is a limited role for local government.
2. **Devolved governance** systems where there is significant input from locally elected governments.
3. **Polycentric** systems with multiple intermediary bodies independent of government.

A fourth, less tangible tendency involves seeing the whole system as a collective actor and forum. However, as noted above, this does not fall under all definitions of accountability.

Measures and consequences

Most systems combine measures of inputs, processes and outcomes.

- **Inputs** – such as compliance with legal obligations, teacher certification and safeguarding are often monitored by requesting reports to the ministry or an authority, or through a management information system.
- **Processes** such as the quality of teaching tend to be monitored by inspectorates (though inspection can play a role in all three types of measurement).
- **Outcomes** are often measured via exams and other assessments (often as part of performance-based accountability' in which incentives are linked to measured performance)[8], but other outcomes can also be monitored too.

System level performance is often monitored via sample or thematic inspections and assessments. Some jurisdictions have a strong focus on this approach and have established bodies to review findings and propose adjustments to policy and practice.

Typical accountability consequences are:

- Additional support or funding
- Re-inspection/increased frequency of inspection
- Warning letters, improvement plans, and category designations.

Some systems also make provision for changes of leadership or governance, and potential funding consequences or fines.

The extent to which governments set their systems up to generate 'market consequences' varies significantly between countries. Such policies can include parental choice, publication of school-level exam and inspection results, and per capita funding.

Developing and reviewing accountability policy

The ministry typically initiates changes in policy and then works with accountability bodies and/or advisory groups to develop detail. Some advisory groups are convened specifically to oversee a programme of reform but some jurisdictions have established more permanent bodies.

International bodies – notably the OECD, have considerable influence over accountability policy. Sometimes this is a result of 'Pisa shock' – when results from international benchmarking tests reveal disappointing results, but many governments also commission reviews from the OECD.

Major changes tend to result from shifts in the government's education strategy. More iterative changes often follow from ongoing monitoring. Ministries tend to be less active in these 'course adjustments', particularly when accountability bodies are independent of the ministry.

Jurisdictions can tentatively be put into four groups based on their approach to making accountability policy:

1. Ministry dominated
2. Shared ministry-accountability body
3. Significant independence or discretion for accountability body
4. Local delegation under strategic guidance

Contexts and effects

Accountability systems bear the hallmarks of the context in which they developed including historic legacies, as well as more recent developments layered on top.

Key contextual influences include:

- Beliefs about the role of the state, and where authority should lie;
- Patterns of political centralisation and governance;
- Underpinning values such as consensus, meritocracy and collectivism – along with levels of trust

More recent influences are then layered on top. Key factors include:

- The varying degree to which "new public management' approaches have been adopted;
- Efforts to 'shrink the state';
- Concerns about inclusion, in particular regarding indigenous groups;
- Priorities linked to economic crises and the covid pandemic.

Studies of performance-based accountability – in which test scores are linked to consequences, tend to report positive effects on pupil attainment. However:

- Evidence on how performance-based accountability might affect different groups is mixed.
- Performance-based accountability can also drive a focus on whatever is being measured – which may or may not be desirable.
- Studies do not always distinguish between the effects of market-based, versus hierarchical consequences linked to performance measures.

Market based systems can drive pupils towards high performing schools or create an incentive to improve, potentially ensuring more pupils are taught in effective schools. However:

- Schools branded underachieving due to their test scores – or based on an external judgement – can lose resources or find it harder to recruit teachers, making it harder to improve.

Horizontal professional forms of accountability are often associated with less compliance orientated and competitive behaviour. Teachers and leaders tend to report that collaborative, peer-to-peer arrangements prompt reflection, problem solving, learning and collective responsibility.

Accountability's effects can be understood with reference to a series of hypothetical 'context mechanism output' configurations:[3]

1. Pressure from accountability systems can lead organisations within the same sector to copy one another and become similar in shape, structure or form. This is more likely to happen under market-based systems; where there is extensive (vertical) teacher-level accountability; and in polycentric or chain of command systems that focus on sanctions.
2. Teachers and leaders may collectively define professional norms, leading to a degree of convergence but through non-coercive mechanisms. This might be driven by professional training, certification, shared standards and interaction through horizontal professional networks.
3. Educational institutions might abandon or adopt practices in order to survive in the face of competitive pressure. This mechanism is largely driven by market pressures.
4. Individuals may be intrinsically motivated to improve in contexts where their basic psychological needs for autonomy, competence and relatedness are met.
5. Institutions may innovate and learn through collaboration and reflection. This can happen in a range of systems but is may be more likely when:
    - ⊄ There is an absence of pressure for immediate improvement.
    - ⊄ School or district inspectors support learning and improvement through context-sensitive professional conversations.
    - ⊄ Resources are allocated for improvement.
    - ⊄ Accountability aligns with professional capacity as well as teachers', schools' and the wider community's values.

[3] Ehren (2025)

Summary References

NB: References for policy documents and additional sources can be found in the full and underlying reports and tables in the CES library on the CES website.

Atkinson, P., Coffey, A., & Delamont, S. (2003). *Key themes in qualitative research: Continuities and changes*. Rowman Altamira.

Burgess, S. (2025). *A review of quantitative causal research evidence on school accountability: updating the evidence*.

Department of Education and Skills (DES). (1984). *Parental Influence at School: A new framework for school government in England and Wales* (Cmnd 9242). Her Majesty's Stationery Office.

Ehren, M. (2024). *Conceptualising "accountability" for the comparative review.*

Ehren, M. (2025). *Context, mechanisms and outcomes of accountability systems in fourteen education systems. Working paper for the Centre for Education Systems*.

George, A. L., & Bennett, A. (2005). *Case studies and theory development in the social sciences*. Belfer Center for Science and International Affairs.

Karsten, S. (1999). Neoliberal Education Reform in The Netherlands. In *Education* (Vol. 35, Issue 3).

Kogan, M. (1986). *Education accountability: An analytic overview*. Routledge.

Majone, G. (1997). From the positive to the regulatory state: Causes and consequences of changes in the mode of governance. *Journal of Public Policy*, *17*(2), 139–167. https://doi.org/10.1017/s0143814x00003524

Maroy, C., & Voisin, A. (2017). *UNESCO: think piece on accountability: background paper prepared for the 2017/8 Global education monitoring report.* https://doi.org/10.13140/RG.2.2.17900.18569

O'Mara-Eves, A., Vigurs, C., Ko, S. Y. (Jessica), & Simon, A. (2025). *Rapid systematic evidence review of school accountability policies in secondary education across fourteen jurisdictions*. EPPI Centre.

O'Mara-Eves, A., Simon, A., Anand, K., Stansfield, C., Vigurs, C., & Ko, S. Y. (Jessica). (2025b). *Technical Report for Accountability Policy in Secondary Education. EPPI Centre.*

Pawson, R. (2024). *How to think like a realist: A methodology for social science*. Edward Elgar Publishing.

Seawright, J., & Gerring, J. (2008). Case selection techniques in case study research: A menu of qualitative and quantitative options. *Political Research Quarterly*, *61*(2), 294–308. https://doi.org/10.1177/1065912907313077

Stecher, B. M., Camm, F., Damberg, C. L., Hamilton, L. S., Mullen, K. J., Nelson, C., Sorensen, P., Wachs, M., Yoh, A., & Zellman, G. L. (2010). *Toward a Culture of Consequences: Performance-Based Accountability Systems for Public Services. Monograph.* ERIC.

Part 1: Introducing accountability

Origins and tensions

The term accountability is closely linked to the term 'accounting' – in the sense of book keeping[4] and has been traced back to the years following the Norman Conquest of England and the writing of the Domesday book.[5] School inspection also has longstanding origins, and is believed to date back to the Napoleonic era.[6] In the modern era, a belief in the importance of accountability underpins many aspects of public service, as embodied in the 'Nolan Principles' of public life, which state that:

"Holders of public office are accountable for their decisions and actions to the public and must submit themselves to whatever scrutiny is appropriate."[7]

Definitions of accountability vary between those that place the concept within "narrow boundaries" and those that draw broader definitions.[8] Back in 1986, the civil-servant-turned-academic, Maurice Kogan made the case for a narrower definition, arguing that it allows for comparison "by reference to categories which remain stable and which do not change according to the context in which they are being viewed".[9] But for the purposes of this international study– *too* stable a definition that is not sensitive to context is a problem.

Firstly, an overly narrow definition born of one context risks being too rigid to 'travel', particularly given that the term accountability does not exist in many other languages. Secondly, if 'the meaning of a word is its use in the language'[10] then we need to be sensitive to how the term – and adjacent terms (like 'évaluation scolaire' in France) – are used. Thirdly, and perhaps most importantly, if the goal in studying accountability is to support policy makers in making decisions regarding how to govern the school system, then the term needs to encompass all of the major factors that have a bearing on these decisions. If a definition of accountability excludes the publication of a school's exam results because this is not associated with a formal sanction, for example, then this could lead to misleading analysis given that publication will probably have informal consequences that shape how a head teacher leads their school. Therefore, while the definitions set out below guided this enquiry,[11] it has been applied with a degree of flexibility, explicitly highlighting how the 'chameleon of accountability'[12] morphs from context to context. In particular, despite

[4] Bovens, M. (2007). p448
[5] Dubnick, M. J. (2002)
[6] de Grauwe, A. (2007a).
[7] Committee on Standards in Public Life. (1995).
[8] Bovens, M. (2007). p448; Kogan, M. (1986); Mulgan, R. (2000).
[9] Kogan 1986 p28
[10] Wittgenstein, L. (2009). S.43
[11] See also Appendix 2
[12] Sinclair, A. (1995)

initially focusing on schools and groups of schools as the key actors, over the course of this study it became increasingly clear that in many jurisdictions, the key actors are teachers themselves, and in some cases, it even seemed that the system as whole was the most relevant actor.

A working definition

In its most general sense, accountability means giving a justification of what one has done.[13] Such justifications can be given in an informal and non-routinised manner, or through more formalised systems. Our starting point for this study is therefore that school accountability involves educational institutions having an obligation to report on, explain or justify what they have done to one or more sources of authority, drawing on various types of information. This may be done through formal or informal means, but some form of judgement is normally reached, which tends to lead to a consequence.

The informal manner involves simply giving an account, taking responsibility for explaining one's actions in a form of social interaction, while the latter implies being called to give an account by an authority who enforces responsibility on another through more standardised and prescribed systems).[14] However it is worth noting that in market forms of accountability, actors may not have an opportunity to give an explanation or justification before a judgement is made and a consequence precipitated.

The 'information' used as part of accountability can come from a range of sources, and be based on various 'quality concepts'. Sources of information include school inspections, standardised tests and compliance audits. Various organisations and individuals may be involved in gathering this information. Information may be provided in relation to a range of measures, including:

- Educational inputs (e.g. curriculum or school expenditure);
- Processes (e.g. teaching quality);
- Outcomes (pupil wellbeing or educational attainment).

Levels of granularity can also vary and relate to individual teacher performance, school performance, school cluster performance, or system performance. A range of individuals can then make decisions about this information, including school managers, inspectorates, departments of education, and the wider stakeholder community (including parents choosing schools).

At the outset of this study a decision was made to consider all forms of quality concepts, measures and purposes, but to restrict the scope to:

[13] Hoffer, T. B. (2000). Accountability in education. In *Handbook of the Sociology of Education* (pp. 529–543). Springer. p259

[14] Mulgan, R. (2000). 'Accountability': An Ever-Expanding Concept?, Public Administration, 78(3), pp. 555–573

1. Formal accountability, in terms of measures, while retaining an openness to informal or unplanned consequences.
2. School organisations and school groupings as the actor.

However, as noted above, the second parameter was somewhat relaxed over the course of this study, in response to emerging findings.

Dissecting the components

There are numerous ways of comparing and categorising accountability systems.[15] Systems can be compared in relation to:

- **Participants**:[16]
- Subject (actor) based typologies: focusing on who is accountable?
- Initiator (forum) based typologies: focusing on who they are accountable to?

1. **The relationship between participants.** Including whether the relationship is:
    - Internal or external,
    - Vertical, horizontal or diagonal,
    - Uni or multi-directional?[17]

2. **What is measured.** Including the degree of emphasis on:[18]
    - Inputs, norms and regulations
    - Processes
    - Outputs

3. **Consequences**

4. **Overall approaches (or instrument types):** several of the above dimensions (participants, relationships and consequences) can be combined to describe overarching approaches to accountability.

5. **Underpinning ideologies.**

The accountability debate

Accountability is a hotly debated topic and has been for a long time. Back in 1974 England's Department for Education and Science announced the creation of a new "Educational

[15] Smith, W. (2025); See also Appendix 2 'Concept Note'
[16] Bovens, M. (2007).
[17] Hooge, E. et al. (2012).
[18] De Grauwe (2007a); Kogan, (1986). p80

Disadvantage Unit", which would work with an 'Assessment and Performance Unit' to improve identification of educational disadvantage and promote good practice.[19] Concerns rapidly grew that this new source of accountability would undermine teachers' autonomy; create curriculum conformity; drive teaching to the test through "backwash"; and that metrics would fail to adequately take into account context.[20]

Fifty years later the debate on accountability has not lost any of its heat, and many of the same concerns remain. A desire to tackle educational underperformance gives rise to new mechanisms for scrutiny and support. Yet these structures and processes are frequently controversial, particularly when they drift into 'surveillance'. In turn, this prompts calls for less centralised, more collegiate and professional approaches.

Jurisdictions around the world are attempting to solve this conundrum in different ways. By cataloguing the workings of fourteen accountability systems; exploring how these systems interact with their contrasting contexts; and by synthesising the evidence on these systems' effects, this study set out to provide evidence to inform a more constructive debate. It was guided by six main research questions:

RQ1. What is the purpose of accountability?
RQ2. What is the structure of accountability?
RQ3: How is accountability policy made and introduced?
RQ4: How is accountability policy evaluated?
RQ5: How does context shape accountability policy and reform?
RQ6: What is the evidence about the effects of the accountability system?

It focused on fourteen jurisdictions: England, Scotland, Wales, Northern Ireland, the Republic of Ireland, Ontario, New Zealand, Singapore, Japan, Estonia, Finland, Poland, The Netherlands and France.

The overall project included four 'work packages'.

Work Package 1: Led by Alison O'Mara-Eves and Antonia Simon. The research team also included: Kusha Anand, Carol Vigurs and Jessica Ko.

i) A pragmatic review of policy documentation via searches of government and educational body websites, with expert validation.
ii) A rapid systematic review of the research evidence on RQ5 and 6[21]

[19] Her Majesty's Stationery Office (1974)
[20] Kogan 1986, pp 71–72
[21] O'Mara-Eves et al 2025

Work package 2. Led by Simon Burgess[22]

This strand of work focused on the effects of accountability systems. It involved a highly targeted approach to identifying quantitative studies that describe (potentially generalisable) causal effects linked to accountability systems – typically through experimental or quasi experimental approaches.

Work package 3. Led by Melanie Ehren[231]

WP3 brought together RQ5 and 6 by identifying hypothetical 'context-mechanism-outcome' (CMO) configurations through a realist inquiry. These were shared with in-country experts who provided feedback.

Work package 4 (this report): Led by Loic Menzies

This work package involved synthesising findings from the three preceding work packages, along with further cross-case analysis.

Full details of the study's methods can be found in Appendix 1.

[22] Burgess 2025
[23] Ehren (2025)

Part 2: The characteristics of accountability systems

Part Two of this report describes fourteen accountability systems' main characteristics, focusing on:

- The systems' intended purposes and underpinning values, as well as how these are changing (RQ1).
- How the different elements of the fourteen systems are structured (RQ2).

Findings in the following sections draw on a policy review by the EPPI centre.[24]

RQ1: Purpose, values and trends

There is considerable overlap between what different jurisdictions are aiming to achieve through their accountability systems. Most countries refer to:

What most accountability systems are seeking to do

Drive educational improvement or excellence;

Promote fairness, equity or inclusion;

Ensure compliance with government policy and regulation (particularly with regard to safeguarding);

Provide transparency and facilitate stakeholder or public engagement;

Promote well-being (of pupils and/or teachers);

Provide data or evidence to support decision making.

Despite these overlaps, there may be some variation in the degree to which they emphasise different purposes. The following word clouds give an indication of different jurisdictions' tendencies to focus on similar terms in their accountability documentation – whilst emphasising different aspects to differing degrees.[25]

[24] O'Mara-Eves et al. (2025b)

[25] We sought to identify documents that were broadly comparable in terms of their purpose, and have chosen those that were most similar in purpose. However, care should nonetheless be taken in interpreting these images since no two documents are truly comparable and this does not represent a robust linguistic

Accountability policy often combines two types of goals.

- **A focus on minimum standards and the 'basics':** typically through compliance checks, core-subject measures, and more frequent inspections for schools judged 'at risk'.
- **Expansive attempts to promote 'improvement for all':** involving wider educational priorities and continuous-improvement strategies.

	Improvement for everyone	**Thresholds and safety nets**
Netherlands[26]	"The law stipulates that education must achieve certain minimum requirements. The Inspectorate ensures that this basic standard of quality is being guaranteed by school governing boards" "We ensure that the basic quality is in order."	"We encourage school governing boards and schools to pursue their own specific ambitions and to achieve improvements above and beyond the basic standard of quality that is required. The Inspectorate wishes to showcase what is going well in schools, in school governing boards, and across the education system as a whole." "In addition, we encourage governing bodies and schools to pursue their ambitions and achieve improvements above and beyond the level of basic quality."

analysis. Sources were: England (Ofsted, 2022; Department for Education, 2018); New Zealand (ERO, 2021); Northern Ireland (Northern Ireland Audit Office, 2023); Netherlands (Inspectorate of Education, 2025); France (French Council for School Evaluation, 2022); Finland (Finnish Education Evaluation Centre, 2024). The documents from France and Finland were somewhat different and focused more on evaluating the national education system.

[26] Inspectorate of Education (Netherlands) (2025) p13; SICI, (2024) p2

England[27]	"The primary purpose of inspection under this framework is to bring about improvement in education provision."	"Inspection provides assurance to the public and to government that minimum standards of education, skills and childcare are being met; that – where relevant – public money is being spent well; and that arrangements for safeguarding are effective."

There are outlying cases of jurisdictions emphasising distinctive goals: These are often linked to specific national priorities derived from national context. For example:

- In New Zealand, considerations of fairness, equity and inclusion are particularly linked to goals related to biculturalism and Māori rights.
- In Singapore and Estonia there is a distinctive emphasis on different cultural or linguistic groups and strategic national or culturally-specific values such as civic development, 'student centred learning' and 'meritocracy'.

Policy documents often link accountability's purposes to underpinning values.

- A desire to promote diversity, inclusion and fairness is often premised on the fact that these are cherished national values.
- The importance of evidence is often linked to 'data-driven decision making' which is viewed as a foundation for fairness.

Similar purposes can be framed with reference to different values, which can, in turn, feed through to different accountability structures.

- Some jurisdictions link transparency to values of integrity and public scrutiny. This has led Estonia to invest in a national tech platform to facilitate data scrutiny;
- Others frame transparency in civic-democratic terms, emphasising community engagement, and stakeholder input. This may play a role in Ontario, Poland and France's emphasis on participatory boards and councils;
- Other systems see transparency as an enabler of public, or consumer choice. This may play a role in England and the Netherlands efforts to provide comparable information to parents.

Trends

Common trends can be identified within recent and upcoming changes across many countries' accountability systems.

Concerns regarding a perceived 'narrow' focus on academic outcomes (both within accountability systems and education systems more broadly) has led to a broadening of the measures used within accountability systems. This trend seems to have

[27] Ofsted (2023)

contributed to a widespread adoption of teacher and pupil well-being measures or school climate and engagement metrics, which is explored in more detail below.

There has been a widespread move towards more 'risk-based', or 'differentiated' accountability systems. This trend overlaps with moves towards 'smarter' or more data-driven accountability systems and may have been precipitated by financial constraints, alongside a desire to promote school or professional autonomy.

Accountability systems have increasingly emphasised diversity, equity and inclusion. This is particularly evident in Ontario where – amongst other actions – the ministry issued a specific policy memorandum mandating equity policies and requesting the integration of anti-discrimination practices within all board operations.[28]

Many jurisdictions have sought to give stakeholder voice and self-evaluation more weight within their accountability frameworks. For example, in France, the 2020 framework mentioned stakeholder participation but did not explicitly set out any required mechanisms, whereas the 2023 framework introduced mandatory involvement of all community members (students, parents, staff).

Recent accountability reforms have often prioritised workforce retention, professional learning and minimising workload. This has had implications for the process of accountability.

- Teacher recertification requirements have been cut back in New Zealand.
- Routine school inspection has been terminated in Estonia.
- There has been a reduction in self-evaluation requirements in England (in direct contrast to most other jurisdictions).

Because accountability systems are an important policy lever, they are often used to drive forward specific reforms and to create alignment – or compliance – with policy.[29]

- In Estonia, a focus on integrating the Russian speaking population, and a decision to unify the language of instruction has been accompanied by an action plan providing tools to monitor quality for pupils for whom Estonian is not their mother tongue as well as sanctions (fines) for schools found not to be using Estonian.
- A heightened focus on curriculum design in England drove a shift in emphasis within the national inspection framework, as well as changes to inspection practice.

Some jurisdictions have sought to make accountability data and information increasingly public, whereas others have gone in the opposite direction.

[28] Ontario Ministry of Education (2013)

[29] Dubnick (2005) notes that "account giving" is often designed to promote "congruence" between organisational goals and a unit or individual's actions p 384–385

- Estonia launched Education Eye / Haridussilm in 2004 to provide the public with "comprehensive information about education establishments, students, teachers and curricula".[30]
- France's 'AMDAC' was established in 2021 to oversee open data and enhance transparency and data accessibility.[31]
- The Netherlands' "Vensters voor Verantwoording" (Windows for Accountability) platform was made available to the public in 2010.
- In contrast, Wales and Singapore have either reduced the availability of information or lessened its granularity, in order to 'lower the stakes'.

As the next section shows, the structure of different accountability systems bears the hallmarks of these various purposes and trends.

RQ2: The structure of accountability systems

This section begins by setting out the key components of the fourteen reviewed accountability systems, including participants, measures (including the tools used to measure), and consequences. It then turns to a broader comparison of systems, before introducing seven common approaches that capture these system's distinctive characteristics and commonalities.

The components

Participants and relationships

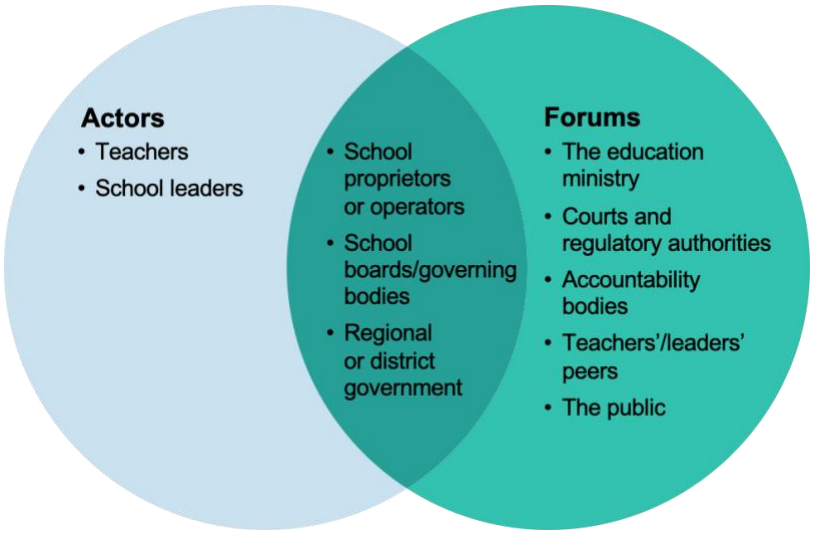
The participants in accountability systems can be categorised as:

1. Teachers;
2. School leaders;
3. School boards or governing bodies;
4. Regional or district government;
5. School proprietors or operators;
6. The courts and regulatory authorities;
7. Accountability bodies (including the inspectorate or teacher certification body);
8. The education ministry;
9. Professionals (including teachers and leaders and their peers);
10. The public.

These participants do not divide neatly between 'actors' (those who are held to account) and 'forums' (those who hold others to account).[32] For example a school board or governing

[30] European Commission (2022), p71
[31] Poisson and Jorgoni (2025)
[32] See Concept note – Appendix 2

body might be responsible for holding a headteacher to account, while also being held accountable by the ministry.



The eight groups of participants are linked together through relationships that can be characterised as:

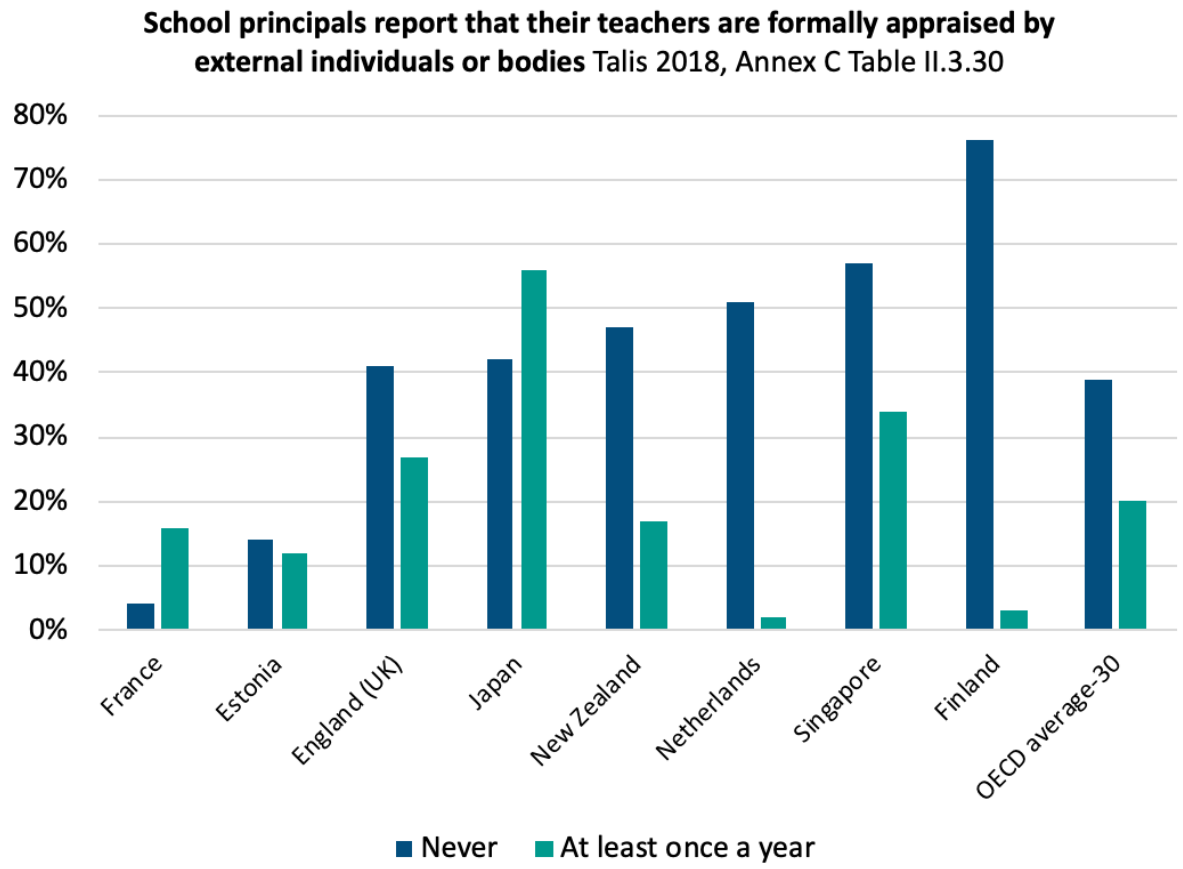
- **Vertical**: where the relationship is hierarchical;
- **Horizontal:** where the relationship is between equals. We use this term to refer to relationships between professionals, or between professionals and the public.

The term 'diagonal' is also sometimes used to refer to relations with an intermediary body, such as an inspectorate.[33]

In some countries teachers are directly held to account by inspectors, the ministry, a teacher certification/professional body, or a network of peers. However in most countries, teachers are held accountable by their managers (school leaders), and managerial relations within a school are not within the scope of this study.

The degree to which teachers are held accountable by bodies external to the school varies considerably. In Finland, more than three-quarters of teachers say they are *never* held to account by anyone external to the school, whereas in France this figure is only 4%, and in Japan more than half are held to account by external bodies at least once a year.[34]

[33] Maroy and Voisin, (2017)
[34] OECD (2018) Data is available for England, Estonia, Finland, France, Japan, Netherlands, New Zealand, Singapore



Data for other jurisdictions unavailable

In some systems school governing bodies are responsible for holding school leaders to account, but in other systems this role is fulfilled by other participants (such as the proprietor, local government, or delegated authority from the ministry). Where governing bodies have significant responsibility as a forum, this tends to be combined with a role as an actor, since the governing body normally then becomes responsible for the school's performance.

The next 'layer' of participants is constituted in various ways in different jurisdictions but these participants are normally both actors and forums.

- **In government-dominated systems, such as France, Poland and Singapore, the next layer tends to be made up of regional or district managers appointed by the central ministry.** Accountability then normally flows upwards to the ministry via managerial relations. France draws its 'recteurs' (who are senior civil servants) from universities, as part of a government controlled-process, with appointments made by the President of the Republic. New Zealand also has regional offices that play some role in supporting schools but there are mixed views on whether this constitutes an accountability role. In England the Department for Education appoints Regional Directors who provide an element of this approach.

- **Some jurisdictions delegate responsibility for education to local government through local authorities or municipalities.** This can be seen in Wales, Scotland and Japan. Accountability can then flow directly to the electorate as well as to central government, leading to a sometimes uneasy relationship.
- **A range of different institutions fall into the 'multi-school proprietor or operator' grouping.** The main types are summarised below:

	Legal / structural model	**Countries & examples**	**Accountability route** *(main elements)*
Formal multi-school legal proprietors	Independent corporate entities that legally own and operate multiple schools, and receive state funding directly.	**Netherlands** (school boards); Ireland (patrons); England (academy trusts).	Inspectorate / ministry / regulator.
Public multi-school operators with statutory boards	Publicly constituted corporate bodies that operate groups of schools under elected governance.	**Ontario** (school boards).	Electorate and ministry.
Quasi-proprietorial but LA-maintained	Foundation or faith bodies own school buildings and appoint governors, but schools remain maintained and funded by local authorities.	**England** (voluntary-aided schools).	Inspectorate and LA for finance and compliance; foundation for ethos and governance.
Non-proprietorial managing authority	Oversees and supports schools but does not legally own them.	**NI** (Education Authority and Council for Catholic Maintained Schools).	Inspectorate / Department of Education.

School boards in Ontario are unusual within CES' jurisdictions, though they are more common in other North American systems. These are directly elected and have responsibility for different subsystems running either English or French schools, secular or religious schools, in overlapping geographical areas. Although these are known as 'school boards' they are categorised as 'school operating groups' in this study, to distinguish them from individual school boards and governing bodies, since this better reflects their role and status.

Courts, regulatory authorities and ombudsmen may only interact with a small number of schools, districts or professionals each year, but they are still part of the accountability mix. This is because they act as a 'safety net' in cases of malpractice or illegality. Moreover, accountability is not always about being called to account, it can also be

about the *possibility* of being called to account, in the case of wrongdoing and exert a passive influence in this way[35].

In some cases, inspectorates, teacher certification bodies and regulatory authorities are part of the ministry, but in other cases they are independent or quasi-independent. Countries where these bodies are independent often also have school operators or proprietors, which results in multiple, somewhat independent centres of power and authority. This can be termed 'polycentricism'. In some cases, independent or quasi-independent bodies do not have hierarchical power to act on their findings themselves. Instead they act as intermediaries, providing information to the ministry, and leaving it to take action. This has been described as 'diagonal accountability'[36].

Teachers' and leaders' peers can also act as a forum, resulting in teachers and leaders combining a role as an actor with a role as a forum. Many systems have structures designed to facilitate these interactions. For example, Singapore has Professional Learning Communities and Network Learning Communities; and New Zealand has Kāhui Ako. Teachers can also have a say in each other's performance appraisals in many jurisdictions. It is sometimes difficult to distinguish between structures that have an accountability role and those whose role is limited to professional development.

In many countries the education ministry interacts directly with schools or teachers as an accountability forum. As part of this role it can make decisions about whether a school requires an intervention or support. It can also be directly involved in decisions about career progression where it has direct authority over teachers.

There are two ways the public can play a role as a forum. Firstly, community members including parents and guardians can hold the school to account through participatory mechanisms, such as allocated places on boards. Secondly, they can form a judgement on a school and 'vote with their feet' where school choice is available.

A different type of actor that does not naturally fall within the original scope of this study is the education system as a collective whole. This is somewhat intangible and involves 'the whole being accountable to the whole'. This idea falls well outside the definition of accountability and parameters set out at the beginning of the study. However, it seems to be an important feature of how accountability is understood in Finland and elements of this

[35] Mulgan (2000), argues that "the effectiveness of accountability mechanisms is to be observed not simply in the occasions when officials are actually brought to account. Much more important in securing compliance is the ever-present threat of being called to account, the potential implicit in accountability. p567

[36] Maroy, C., & Voisin, A. (2017).

approach can be found in other systems. This approach is therefore explored in more detail in 'overarching findings'.

An initial typology of accountability can therefore be described, based purely on who participates. The table below sets out some typical cases of four approaches within this typology. However other jurisdictions do not fall so clearly within a single category. Instead they hybridise between approaches. Even in the case of England, which is broadly polycentric, this is layered on a previous local government-centred system which has in some ways now moved towards centralisation (by giving power to directly appointed regional teams, and by subsuming the Education and Skills Funding Agency and Teacher Regulation Authority into the ministry).

Highly centralised	***Polycentric***	***Local governance***	***Collectivist***
Accountability is to the ministry or its appointed representatives. Limited role for local government. Few non-governmental bodies. *Examples of typical cases:* Singapore, France, Poland	Multiple intermediary bodies (actors and forums) independent of government *Examples of typical cases:* England, Netherlands	Significant input from locally elected government *Examples of typical cases:* Scotland, Wales, Japan, Estonia	System level responsibility *Examples of typical cases:* Finland

Quality measures and tools

Accountability measures can be qualitative, quantitative, or compliance based, and systems can focus on:

- **Inputs, norms and regulations.**
- **Instructional and organisational processes.**
- **Outputs, including both short and long-term results.**

Jurisdictions invariably combine elements of all three of these 'quality concepts'.[37]

Inputs, norms and regulations

Accountability for inputs usually involves management information systems, reports, audits or aspects of inspections that check for compliance. The aim is often to avoid malpractice and the most commonly emphasised dimension in this area is safeguarding. Checks on teacher qualifications/certification are common too. Where input monitoring takes place separately to inspection, actors are typically required to submit information directly to the ministry or to a local/regional manager.

[37] See appendix 2

Schools and teachers are also often required to provide information on provision for pupil equity, as well as teachers' participation in professional development. Input-focused measures of equity involve reports on what support has been put in place for socio-economically disadvantaged pupils or those with special educational needs. These reporting requirements are often combined with external, outcome-based monitoring, but this is harder in the case of special educational needs given the level of variation in need level and challenges with identification (see CES SEND report).

Spotlight on SEND accountability in Northern Ireland and the Republic of Ireland

In Northern Ireland, the Special Educational Needs and Disability Act (Northern Ireland) 2016 imposes legal obligations on Boards of Governors and the Education Authority to provide for a child or a young person with SEN. These obligations include the appointment of a Learning Support Coordinator in each school and personal learning plans for pupils with SEN. All schools must produce, monitor, and review an accessibility plan on three equality aspects. These aspects are 1) access to the curriculum, 2) improving the physical environment, and 3) the provision of information in an accessible format (EA, 2024a, p. iii).

In the Republic of Ireland, the DEIS programme (Delivering Equality of Opportunity in Schools) provides targeted support for schools serving communities with high levels of educational disadvantage. Approximately 30% of schools are currently involved. All DEIS schools are required to develop a three-year DEIS Action Plan for Improvement, and this must be reviewed annually. Plans must be data-informed and focused on the effective use and impact of DEIS resources.

The Department's Inspectorate, evaluates DEIS Action Planning through school inspections and thematic reports. Additional supports under DEIS, such as Home School Community Liaison officers and School Completion Programme projects, are monitored via annual reports to the Tusla Education Support Service (TESS). Additionally, the National Council for Special Education (NCSE) Inclusive Education Framework provides schools with a framework for self-reflection as well as a qualitative measures to document and evaluate inclusive practices (National Council for Special Education, 2011, pp. 11–14).

Input-monitoring can also take place through inspection. For example inspectors may check teachers' certification, or review compliance with pupil safeguarding requirements (though safeguarding can also involve a deeper focus on 'process').

Processes

This 'quality concept' focuses on features such as teaching and learning, school climate, leadership and decision-making.[38] It is particularly associated with school inspection, as well as professional and peer-to-peer forms of accountability.

Judgements about processes tend to involve multiple data sources. These can include:

- Quantitative data from parent and student surveys;
- Peer or expert review and self-assessment;
- Qualitative data gathered through interviews, focus groups and lesson observations.

More detailed comparison of inspection processes is provided below

Peer judgement can be conceived of as a 'measure' of teaching quality. This is rarely quantitative or formal within learning-focused networks, but peers do sometimes have input into each others' performance appraisals.

School climate is an increasingly prominent 'quality concept' and is monitored in a number of ways, including through standardised well-being measures. As was noted in earlier, concerns regarding teacher and pupil wellbeing have contributed to a growing emphasis on workload, bullying, safety and the promotion of mental health across most CES jurisdictions.

England and Wales	State mandated, standardised pre-inspection survey of staff and pupils (+parents and governors in Wales) on perceptions of bullying, behaviour, safety and wellbeing. In England, live data is published from parents on similar questions through the 'Parent View' platform.
Scotland	The Scottish Government provides data infrastructure for local authorities to collect data on health and well-being on all children from P5 to S6 (approximately age 8.5 to 18)
Ontario	School climate surveys are conducted biennially with students, staff, and parents to assess perceptions of safety and inclusion. Data from these measures inform targeted interventions (e.g., anti-racism training, and mental health supports (OME, 2014; OME, 2024),
Estonia	The Estonian Education Information System (EHIS) compiles information on indicators such as student satisfaction, pupil relationships, bullying rates. This data can inform risk-based inspections.
Finland	Sample based evaluations of school climate include student/parent/teacher surveys and qualitative data
Netherlands	Schools required to monitor pupils feelings of well-being at least once a year using a standardized instrument

[38] De Grauwe (2007a)

Outputs

Output or outcome-focused 'quality concepts' are typically – but not solely – associated with 'performance-based' or 'results-based' accountability[39]. Outcome measures usually involve assessments of pupils' academic attainment or progress,[40] but statistics on attendance; exclusion and drop out rates; teacher turnover (Estonia); and the number of violent incidents (Ontario) are also used.

Some countries use longer-term 'destination' measures such as progression to future schooling and employment.

- Estonia monitors employment rates of recent graduates,
- Finland's emphasis on system monitoring involves using longitudinal sample-based studies to monitor graduation and employment rates, particularly from vocational education.
- In England, 'destination data' is published as a school-level metric, but this focuses on broad measures of sustained employment or education.

In some countries, pupil attainment is contextualised to take into account the school's circumstances and/or intake, and some measures focus on progress rather than attainment. In France, the Directorate for Evaluation, Forecasting and Performance (DEPP) produces an 'IVAL' metric which measures the value added by a 'lycée' or 'collège' and attempts to 'neutralise' contextual effects.[41] Data is publicly available and is intended to inform school evaluation and improvement. In contrast, England has moved away from its previous contextual value-added measure, although it retains a focus on progress. Inspection can also play a role in contextualising data.

[39] De Grauwe (2007a)

[40] The characteristics of different countries' assessment systems – for example the use of test based, or internally assessed metrics – is covered in CES' forthcoming report on assessment.

[41] Direction de l'évaluation, de la prospective et de la performance. (2025, September). *Indicateurs de valeur ajoutée des lycées*. Direction de l'évaluation, de La Prospective et de La Performance (DEPP). https://catalogue.depp.education.fr/index.php/catalog/274

	Inspection	Quantitative Attainment measures
Progress	**Wales:** External inspections evaluate schools using indicators that include learner progress. **Scotland:** National Improvement Framework (NIF) data, tracks progress on literacy, numeracy. **Finland:** Longitudinal evaluations track student progress over several grades to monitor trends in competence development.	**England:** Progress 8 Measure. **France:** IVAL school metrics are based on pupil starting points.
Context	**Scotland:** Contextualized evaluations consider factors like deprivation indices.	**France:** IVAL school metrics take into account school context/pupil demographics. **Scotland:** Destination data, disaggregated by key equity characteristics. **England:** School-level performance measures disaggregated by demographic groups.
Long term outcome measures *(often system level, rather than school level)*		**Scotland:** Destination data, disaggregated by equity characteristics, used to assess how well schools and the wider system are preparing young people for life beyond school (not at school level). **Ontario:** School boards publicly report on graduation rates, credit accumulation, and the effectiveness of career education initiatives. **Estonia:** Estonian Education Information System (EHIS) includes information on employment rates of recent graduates (not at school level). **Poland:** New system tracks the career outcomes of secondary school graduates (not at school level). **England:** School level destination data published alongside attainment/progress.

The extent to which school-level data is made available to the public varies. Eight out of fourteen countries proactively publish school-level data, but in other countries, the media may compile league tables (for example by requesting data directly from schools or submitting freedom of information requests). Both approaches might have significant consequences – intended or unintended.

Public reporting of school-level outcomes by the government is often intended to inform parental choice. It is therefore often a feature of market based approaches to accountability. This has implications for the system's actual and/or perceived consequences (see below). Other countries pursue transparency for democratic rather than market-accountability reasons and some countries have reduced the availability of school-level data (as noted in relation to Accountability Trends, above).

Approaches to inspection

Almost all jurisdictions have some form of inspectorate function, but the way this is structured varies considerably. Inspectorates typically focus on individual schools, but as noted earlier, sometimes they also focus on teachers (as in France).

Not all countries have a systematic national approach to school-level inspection.

- Finland abolished external inspection in the early 1990s, and at present FINEEC conducts school evaluations as part of a broader role that aligns with the systems' collectivist orientation.
- In Ontario, oversight happens through school boards, self-evaluation, and the Teacher/Principal Performance Appraisal processes.
- In Japan, schools must publish their self-evaluation and municipalities and prefectures have flexibility over inspection – however there are national guidelines.
- In Poland, the ministry's regional leads have considerable discretion over their supervision plan.
- In France, the 2019 'Loi pour une école de la confiance' sought to create more national coherence to inspection (see 'Development and Evaluation').

The domains that are reviewed as part of inspection processes are similar across most jurisdictions. Although there differences in terminology, particularly in relation to assessment or curriculum, similar facets may be nested within specifications of 'teaching' or 'educational quality'.

It is beyond the scope of this report to delve into the detail of each inspection framework, but the main domains tend to be:

- Leadership and management
- Quality of teaching and/or learning
- Outcomes
- Pastoral care, well-being, and/or personal development
- School learning environment/climate/safety

England is unusual in including pupil behaviour as one of its main areas – although this is likely to fall within a subdomain of many jurisdictions' definitions of safety and school climate. Scotland's explicit reference to equality and inclusion is unusual too.

England	Quality of education, behaviour & attitudes, personal development, leadership & management
Wales	Learner progress; Wellbeing; Teaching/learning experiences; Leadership; Care/support
Northern Ireland	Leadership/management; Quality of provision; Outcomes for learners
Scotland	Can vary, but include: Leadership of change; Learning, teaching and assessment; Raising attainment and achievement; Ensuring wellbeing, equality and inclusion
Ireland	Leadership, teaching, curriculum, school planning, student/parent feedback.
Ontario	Focus is on self evaluations using the School Effectiveness Framework, and on School Board level accountability. School inspections private sector only
New Zealand	Quality of education, teaching, leadership, curriculum, learner success & wellbeing
Singapore	Customised validation based on School Excellence Model
Japan	Exemplar guidelines only: Curriculum and teaching, pastoral guidance, career guidance, safety management, health management, special education, school management and organisation, professional development, cooperation with parents and local residents, premises and facilities
Finland	Inspections abolished in the early 1990s (FINEEC continues to conduct evaluations as part of its system-level functions)
Estonia	Compliance with regulations, areas for improvement
Poland	Legislation no longer defines specific criteria or aspects to be evaluated in external inspections focusing on educational processes and / or the effectiveness or outcomes. Inspections focus on compliance and can address any aspect within the scope of pedagogical supervision. Inspection sheets, including criteria, are published by the Minister each year
The Netherlands	Teaching-learning process; Secure environment and atmosphere; Learning outcomes; Management, quality assurance and ambition
France	Pupils' learning and educational pathways, and teaching; pupils' life in school, their well-being, and the school climate; the actors, strategy, and functioning of the establishment; the establishment within its institutional and partnership environment

Inspections in most – but not all countries – are graded, either on an overall basis, or by inspection domain. Countries including Ireland, Wales and New Zealand do not grade but make narrative reports publicly available. However:

- Ireland still provides a 'quality continuum', with descriptors to be used as part of narrative reports.
- In New Zealand, decisions by the inspectorate to make a 'return visit' are sometimes interpreted as a form of grade.[42]

Almost all countries make inspection reports publicly available. Exceptions include Singapore and France. In Japan there is no systematic publication of external evaluations but schools must publish their self-evaluation.

[42] Udahemuka, M. (2016).

Most countries operate some form of differentiated, proportionate, or risk-based approach to inspection, such that the frequency of school inspection varies. In some cases, this involves a rapid return to schools considered a cause for concern (as in New Zealand, the Netherlands and England). Other countries like Estonia and Ireland take a data-informed approach to risk identification, often combining this with a programme of thematic evaluations.

Spotlight on risk based inspection in Estonia

Estonia's risk-based inspection model was introduced in 2014, and replaced cyclical evaluations with a system in which inspections are triggered when risk indicators signal potential underperformance. As a result, Estonia is moving toward a model in which inspections concentrate on a small set of low-performing schools

Risk identification relies on systematic data use. The Ministry of Education and Research draws on the Estonian Education Information System (EHIS), which compiles indicators such as dropout rates, student satisfaction, and assessment results. Schools showing concerning patterns are prioritised for inspection.

Thematic evaluations follow similar criteria: EHIS data and well-being surveys help identify systemic issues affecting particular groups of students. Evidence from thematic inspections lead to policy adjustments, and refinements have been made to the risk-identification process to ensure contextual sensitivity.

On this basis, we can describe five broad groups of inspection systems:

Group 1: *Regular inspections lead to publicly available, graded reports**	England, Northern Ireland, Scotland, Netherlands
Group 2: *Regular inspections lead to publicly available narrative-focused reports**	Wales, Republic of Ireland, New Zealand**
Group 3: *Regular inspections lead to reports that are not publicly available*	France, Singapore
Group 4: *Risk-based, ministry-directed or locally determined inspections with no public report*	Estonia, Poland, Japan
***Group 5:** No inspection*	Ontario, Finland,

*Inspection frequency may vary by characteristics or by previous grade

** Inspections may result in a decision to make a return visit and this 'return' may be interpreted as a form of grade

Self-evaluation is a feature of most inspection systems. Most countries have given self-evaluation increased importance in recent years (as noted earlier). Countries typically provide frameworks and tools to support this process and self-evaluation is normally combined with validation or oversight which might come from the inspectorate itself and/or the local authority.

- In France the Council for School Evaluation sets out a detailed process and provides toolkits containing materials such as questionnaires for all key stakeholders.
- In contrast, in England, schools are still required to self evaluate, but the previous self-evaluation framework was abolished with the intention of reducing paperwork and workload.

Approaches to professional accountability

School leaders – including principals, other members of the management team and line managers – are normally responsible for professional oversight of their teachers. These internal, managerial relations are beyond this report's scope.

There are two main forms of vertical, teacher-level accountability beyond internal managerial relations. The first focuses on *compliance*, particularly cases of professional malpractice. The second is more developmental and focuses on professional *practice*.

Countries are split between entrusting vertical, compliance-focused accountability to a distinct teacher registration body/council, versus managing this as part of ministry and delegated structures. Anglophone jurisdictions generally opt for an independent or quasi-independent body, in keeping with their tendency towards more polycentric governance.

Some countries also use system-level, vertical structures to hold teachers to account for their professional practice.

- In Ontario and Singapore, state or province mandated systems lead to judgements with significance beyond internal school management. However, school leaders still play a role in these processes.
- The teaching councils' roles in Scotland and New Zealand extend to professional practice.
- In France, the inspectorate directly evaluates individual teachers.

'Horizontal', peer-to-peer and network-based structures often play a role in professional accountability. These structures can involve peers within an individual's school (internal), or they can extend beyond a single school through networks. Networks tend to be somewhat adhoc or localised, often featuring as part of school improvement initiatives in areas facing particular challenges (for example Ireland's DEIS programme and France's Réseaux d'Education Prioritaire).

- Singapore's school clusters provide a system-wide structure for professional networks
- Japan has long been known for its distinctive 'Lesson Study' system, which might be argued to serve a professional accountability role, alongside professional development.

In the table below, we have listed a range of initiatives that may potentially play an accountability role.[43]

[43] As noted in Part 1 'accountability' is a fuzzy edged concept and these difficulties typify the challenges with overly rigid definitions of the term in a comparative context.

RQ2: Teacher accountability

	Home countries				Anglophone countries			Rest of the world		Europe				
	Eng	**Wales**	**NI**	**Scotland**	**Ireland**	**Ontario**	**NZ**	**Sing**	**Japan**	**Finland**	**Estonia**	**Poland**	**Neth**	**France**
Vertical teacher accountability – compliance focused *(outside of day-to-day school management)*	TRA (Exec agency within ministry)	EWC	GTCNI (currently undergoing reform)	GTCS	Teaching Council	Ontario College of Teachers	Teaching Council of Aotearoa NZ	Ministry	Prefecture	No national evaluation /registration of teachers	No separate structure outside of ministry. Embedded in employment law	No separate structure outside of ministry. Embedded in employment law / Teachers' Charter	Board/ ministry based? Teacher registration legislation proposed but appears not to have been introduced.	Malpractice dealt with through civil service systems
Vertical teacher accountability – developmental *(outside of day-to-day school management)*				Y		Y (TPA conducted by Principals but under a province-mandated system. Leads to a formal rating)	Y though scaled back	Y EPMS	Licence renewal abolished in 2022. Schemes for deployment of highly effective teachers			Structured process for managing evaluation and rating		Y – teacher level inspection with regular reviews
Horizontal professional *(may not be system-wide and could be voluntary, informal or CPD focused)*	Can take place through LAs/MATs, Teaching school hubs	Previously Regional school improvement consortia	Collaborative Area Learning Communities – not explicitly accountability function	Regional Improvement Collaboratives – being wound down	Teacher Professional Networks, and within DEIS	System requirement to support 'collaborative professionalism'	Kāhui Ako – variable and not explicitly accountability focused	PLCs and NLCs	Lesson study an established feature of system	Peer to peer accountability described as key. Networks for professional learning	Collaborative networks may play a role in some school improvement programmes	Promotion and evaluation involve peer review. Training hub network	VO-Raad and PO RAAD networks for school boards	Possible feature of REP

System-level accountability

Accountability often exists at a 'system level'. All jurisdictions use information from accountability in order to gauge the education system's performance, and to inform system-level improvement. However some jurisdictions have accountability structures specifically designed to fulfil this function.

Most systems use sample based surveys to monitor wellbeing, workload and satisfaction. Thematic and sample-based inspections are also common, and some systems also run sample-based tests.

England	The National Reference Test is administered annually and shows if student performance in English and maths at GCSE level has changed from year-to-year. It was taken by 13,000 Year 11 students from over 300 schools in 2024.
Ireland	NAMER is a national study of reading and mathematics at primary school level providing a snapshot of the mathematical and reading skills of second and sixth class pupils. It is used to identify areas of strength and weakness, and to inform educational policy and practice.
Ontario	The Pan-Canadian Assessment Program (PCAP) involves cyclical assessments of reading, mathematics, and science across Canadian provinces and territories.
New Zealand	The National Monitoring Study of Student Achievement (NMSSA) aims to assess and understand student achievement based on a nationally representative sample of students in Years 4 and 8. The focus/curriculum areas vary from year to year. University of Otago also runs Curriculum Insights and Progress Study.
Finland	FINEEC evaluates the attainment of learning goals related to the objectives of pre-primary, primary and lower secondary education core curricula.
Estonia	Sample-based national standardised tests conducted in Years 3, 4, 6 and 7. Results used for national monitoring.
The Netherlands	Peil.education is the set of periodic benchmark surveys in primary, secondary and special education overseen by the inspectorate and providing input for dialogue about the content, quality, and level of education in various subject areas.
France	Cedre is a standardised national sample-based assessment similar to PISA, used to measure pupils' skills at the end of primary schools. A different subject area is covered each year as part of a rolling 6-year cycle. Areas include: Language skills; Modern foreign languages; History-geography and civics; Experimental sciences; and Mathematics.

It might be argued that the term 'accountability' should not be applied at the system-level because there is no distinct 'actor' or 'forum'. There are two responses to this argument, though both sitsomewhat outside of this studies' original scope:

- The whole interlinked system might be viewed as an actor, which is internally accountable in countries with a heavily collectivist orientation (for example in Finland)

- Government and its officials could be characterised as an actor, with the public playing the role of the forum, as part of a political or democratic accountability.[44]

Spotlight on sample based evaluation in Finland

FINEEC conducts sample-based evaluations of learning outcomes in different subjects, which assess the achievement of curriculum objectives (SABER 2012, p. 7). The goal of these assessments is to monitor national trends in student achievement against objectives in the national core curriculum, analysing disparities (e.g., gender, language background, socio-economic status) (FINEEC, 2024, p.17).

The results of sample-based evaluations are not publicly available at the school level. While FINEEC publishes national-level summaries and reports, individual schools and education providers receive feedback on their results, which are compared against the national average. This feedback is intended for internal development purposes and helps schools identify their strengths and areas for improvement (Eurydice, 2025). Parents and other stakeholders can access aggregated national data, but detailed school-level results are not disclosed publicly. This is believed to play a role in maintaining trust within the system and avoiding ranking schools (Eurydice, 2025).

Consequences

When describing the consequences of accountability it is important to note that:

- **It is not just the nature of a consequence, but the frequency with which it is deployed that matters.** A system may have provisions for a school to be closed in response to accountability but – if this seldom happens – it may be less significant than an apparently less 'severe' consequence that is more frequently deployed and feels like a more salient 'threat'.
- **The severity of a consequence is partly subjective.** A loss of reputation in a tight-nit community might be perceived as more severe than a 'hard' consequence.
- **An actors' response to a consequence is contextually determined.** It can depend on factors such as: resourcing and support (which impacts on capacity to change); degree of trust; agreement or alignment between school self perception and the external judgement; inspectors' style of feedback or communication. These factors are explored in detail in the Realist Review,[45] which underpins the analysis in 'Contexts and Effects'.

[44] Bovens, M. (2007). Sinclair, A. (1995).

[45] Ehren (2025)

A continuum of consequences can be described. The most widespread steps in response to accountability measures are:

- Additional support or funding.
- Re-inspection/increased frequency of inspection
- Warning letters, improvement plans, category designations (e.g. 'requires improvement').

Although most countries have resourcing and support arrangements that are closely linked to accountability judgements, this is less marked in some countries than others.

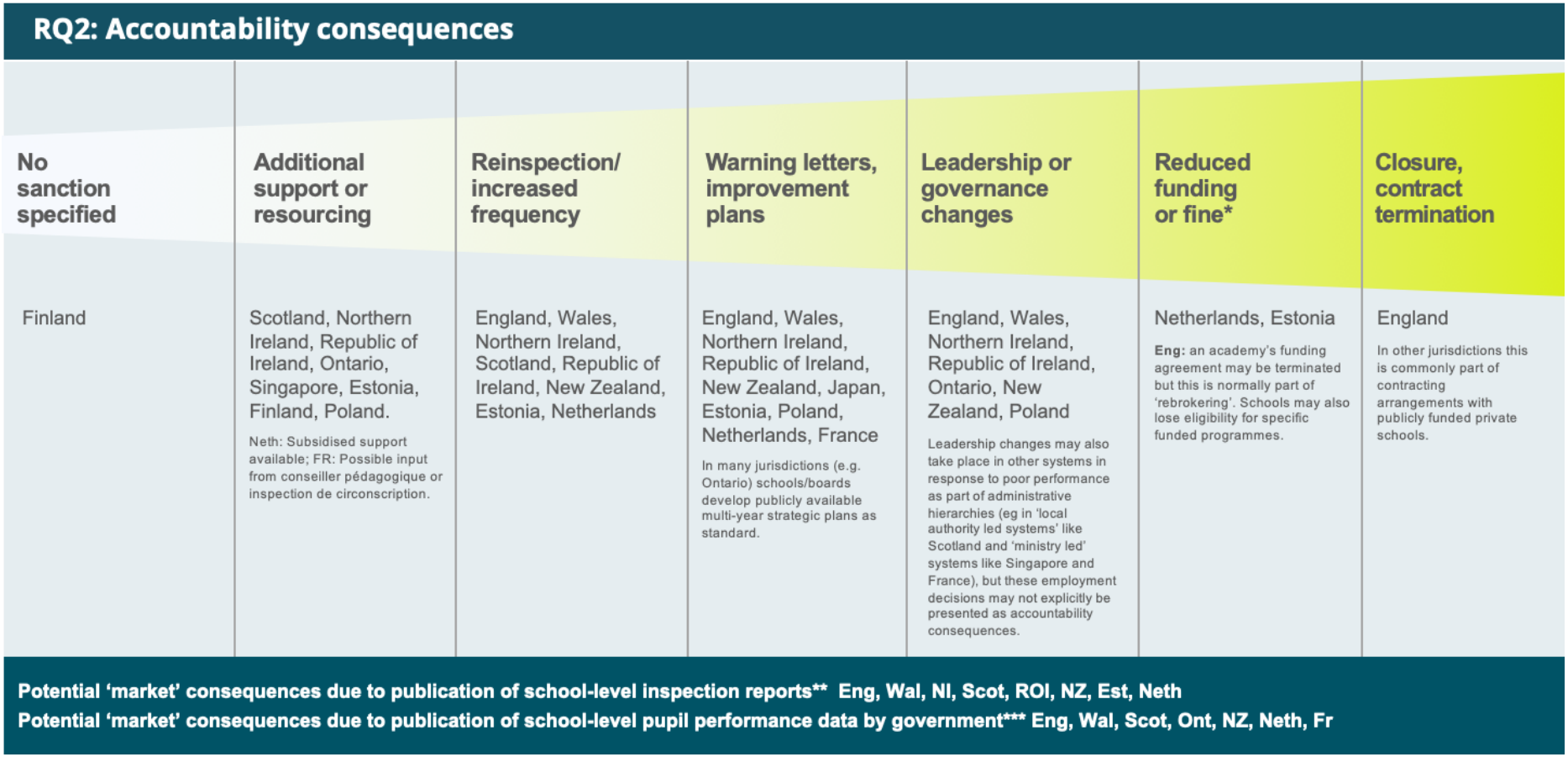
- England appears to be unusual in not having a systematic approach to support or additional resourcing as part of its accountability consequences (until the recent introduction of 'Rise teams'). Instead, the principle behind England's current system is that support should come via a school's management or the responsible body (local authority or multi-academy trust), or that support will follow from a change of governance or leadership.
- No information was found regarding systematic additional support and resourcing in New Zealand, beyond governance level interventions.
- In the Netherlands, the main form of support is subsidised access to resources.
- In France, schools considered to be struggling are likely to receive additional input from their allocated "conseiller pédagogique" or "inspection de circonscription" but this does not appear to be framed as an accountability consequence. Instead, it is a normal part of school management.

Reductions in funding and fines (usually for compliance failures), and school closure are much less widespread.

- Fines are used in Estonia and the Netherlands.
- A school's funding agreement may be ended in England as part of a school being academised or 'rebrokered' (thus this consequence combines a governance change; with funding implications; and the 'closure' of the existing school followed by reopening in a new guise).
- Category designations such as 'special measures' may result in school leadership losing a degree of financial autonomy, as is the case in Wales.
- In systems where private schools operate under government contract, this contract can be ended as part of school accountability. However, such sub-systems tend to represent a small part of the overall system.

These consequences can be broadly described as hierarchical. This is because they are imposed by a 'superior' body, either as part of a 'chain of command', or through a more 'diagonal' relationship with an intermediary body like a regulator or independent inspectorate, in a polycentric system.[9]

RQ2: Accountability consequences



*Refers to actions by a hierarchical authority; market mechanisms may also reduce funding.

**In some systems, schools may choose to publish reports, for example on their websites.

***In some systems, media compile performance measures; practices may also vary locally (e.g., by municipality in Japan).

Most jurisdictions have systems to sanction teachers for serious misconduct and some jurisdictions have systems that reward teachers for their professional practice.

- In Singapore and Japan, teacher evaluations can feed into decisions about redeployment, for example as part of a cadre of expert teachers.
- In New Zealand, the Teaching Council distinguishes between 'conduct' and 'competence' concerns, and Professional Practice Evaluators may assess a teacher's competence against the professional standards in response to concerns.[46]

Accountability consequences can also operate through market mechanisms and where funding follows pupils, this can lead to increases or decreases in funding.

The extent to which schools face market consequences can depend on factors including:

- ∉ policies that promote or facilitate parental choice;
- ∉ publication of school-level performance data and inspection results;
- ∉ the proximity of other schools;
- ∉ arrangements where funding follows pupils.

Pisa provides some data on the extent to which schools experience competition.

- England, Singapore, Japan, Poland and the Netherlands report widespread competition.
- In Japan, competition is concentrated at the upper secondary level, and the availability of choice differs between dense urban areas and rural or remote areas.
- Data is not available for Wales, Scotland, Northern Ireland and Ontario.

No other schools compete for our students
(low scores represent high competition)

Eng	Wal	NI	Scot	ROI	Ont	NZ	Sing	Jap	Fin	Est	Pol	Neth	Fra
8%	N/A	N/A	N/A	13%	N/A	11%	2%	6%	62%	22%	9%	7%	37%

Source: Table II.B1.6.9 Pisa 2022 Table II.B1.6.13 pisa 2022 – figure for England is 'UK wide', but here it is included under 'England'; as England represents the most pupils. The true figure is likely to be higher once other home nations are excluded.

The extent to which the government makes data about school performance or quality available to the public is a potential indicator of the degree to which policy makers *intend* to mobilise market mechanisms.

[46] As noted earlier, teacher-level accountability in New Zealand has been scaled-back in recent years

- England, Scotland, New Zealand and the Netherlands are the only four of the studied jurisdictions in which the government actively and routinely publishes both graded inspection reports and pupil performance data at a school level.[47]
- Even within jurisdictions that publish inspection and performance data, practices can vary. For example, far less detailed information is published on the Scottish government's 'Secondary School Information' website than on the English Department for Education's 'Compare School Performance' website.
- Some jurisdictions explicitly ban or rule out the publication of league tables (Ireland) or inspection reports (France).

[47] In some jurisdictions information is available or can be sought out, and may be used by the media to compile 'league tables' (and schools may choose to publish their inspection reports on their websites). This can have accountability consequences, but suggests that the government may not be actively seeking to drive accountability through market mechanisms.

Summary of accountability consequences

	Home countries				Anglophone countries			Rest of the world		Europe				
	Eng	Wales	NI	Scot	Ireland	Ontario	NZ	Sing	Japan	Finland	Estonia	Poland	Neth	France
Publication of school-level inspection report	✓	✓	✓	✓	✓	N/A	✓	N	N	N/A	✓	N	✓	N
Government publishes school level pupil performance data (1)	✓	✓		✓	N	✓	✓	N	N	N	✓	✓	✓	✓
Teacher level consequences (2)							✓	✓	✓					✓
No sanctions specified										✓				
Additional support or resourcing		?	✓	✓	✓	✓		✓	✓	✓	✓	✓	(3)	(4)
Reinspection / more frequent inspection	✓	✓	✓	✓	✓	N/A	✓				✓		✓	
Warning letters, improvement plans, designation within a category)	✓	✓	✓		✓	(5)	✓		✓		✓	✓	✓	✓
Leadership or governance changes	✓	✓	✓	(6)	✓	✓	✓	(6)				✓		(6)
Reduced funding or fine (7)	✓ (8)										✓		✓	
Closure, ending of contract or 'rebrokering' to another provider (9)	✓													

Notes:
Blanks represent 'none known' rather than a definitive absence
1 In some systems the media may still compile performance measures and practices may vary by area
2 not including employment decisions by head and peer-to-peer support or punishments for severe misconduct
3 Subsidised, inspection-linked support available
4 Possible input from a conseiller pédagogique or inspection de circonscription (1er degré),
5 Boards develop publicly available multi-year strategic plans
6 Potentially as part of an employment decision
7 Termination of funding agreement for academy
8 This refers to actions by a hierarchical authority. Market mechanisms may also lead to reductions in funding
Not including small, publicly funded independent sector

Comparing Systems

Functions

Accountability can be said to have three main functions:[48]

- **Control and evaluation**
- **Improvement and support**
- **Liaison**: acting as a go-between, helping to inform schools of decisions taken by the centre, and informing the centre of the realities at school level

Most systems combine all three functions but the combination of narrative-only inspections and a clear link between accountability and support in some countries suggests a particular emphasis on support and improvement.

- This approach can be found in Ireland, Singapore, Estonia and Poland (while Finland's relatively unique system also focuses on support and improvement).
- In contrast, the absence (until recently) of a systematic link between accountability and support, alongside graded inspections in England, suggests a greater focus on control and evaluation.

Some jurisdictions seem to deploy different elements of their accountability system to fulfil different functions. For example, they might use peer-to-peer and cluster-based professional accountability to deliver 'support and improvement' functions, but pursue 'control and evaluation' through risk-based or compliance-centric inspections targeting schools below a threshold. The question of whether these approaches might be complementary, or whether they lead to tension and incoherence, is discussed further in Overarching Findings.

Although there is little explicit reference to 'liaison', this concept is closely linked to some of the 'system level' accountability approaches discussed earlier. Sample-based and thematic inspections, along with sample-based exams and assessments, play a particularly important role in this function, and this function underpins Finland's approach to collective accountability.

Accountability types

Seven 'types' of accountability have been described in the literature, which bring together multiple dimensions of accountability systems, including participants, relationships and consequences[49].

[48] De Grauwe, (2007b)
[49] West, A. et al. (2011).

Hierarchical Accountability:
Schools are accountable upwards through formal managerial structures to bodies like a local or district authority, the ministry or inspectorate for performance targets, financial management, and other aspects of functioning. Sanctions can include identification as a school "causing concern," heightened scrutiny, replacement of a governing body or a significant reputational consequence.

This plays a significant role in almost all systems, with Finland standing out as an exception.

Market Accountability:
Schools are accountable to consumers (parents) through mechanisms like published examination results ("league tables") and published inspection reports. Weak schools risk reputational damage leading to reduced enrolment, loss of funding, and potential closure due to declining pupil numbers and associated funding cuts.

This appears to be most prevalent in England and the Netherlands, however, as noted earlier, a range of policies can be found in different jurisdictions which may drive some market consequences, and schools in many systems report competition with another school for pupils.

Contract Accountability:
Accountability is contractual, and sanctions can include termination of the contract with the school operator leading to closure or a change of management.

This is often used for publicly funded private schools, however these schools normally only represent a small part of the system. England is unique in deploying contract accountability at a much larger scale. Because it is typically associated with polycentric approaches to accountability it is grouped as part of that approach below.

Legal Accountability:
Schools are held accountable through the legal system for duties, (e.g. staffing, financial management, special educational needs, and health and safety). Sanctions can include financial penalties, criminal sanctions, and tribunal hearings. The analysis within the present study also uses this 'type' to refer to monitoring of compliance with statutory regulations.

This is normally only intended to play a small part of the system, usually when malpractice or non-compliance has occurred. However, many systems also require schools to demonstrate compliance with regulations, or have structures like 'ombudsmans', audit offices and Rights Commissioners that monitor compliance with legal rights and statutory regulations, as can be seen in Ontario.

Professional Accountability:
Teachers are accountable to their peers and professional bodies for conduct, competence, and adherence to standards. Sanctions include warnings, suspensions, or bans from teaching.

This takes multiple forms, as noted earlier. Some form of professional accountability is therefore fairly widespread, but this can include compliance focused accountability related to malpractice, and vertical teacher appraisal, as well as the horizontal peer-to-peer mechanism that are more typically associated with this form of accountability. These rather different forms are separated out in the 'common approaches' presented later in this report.

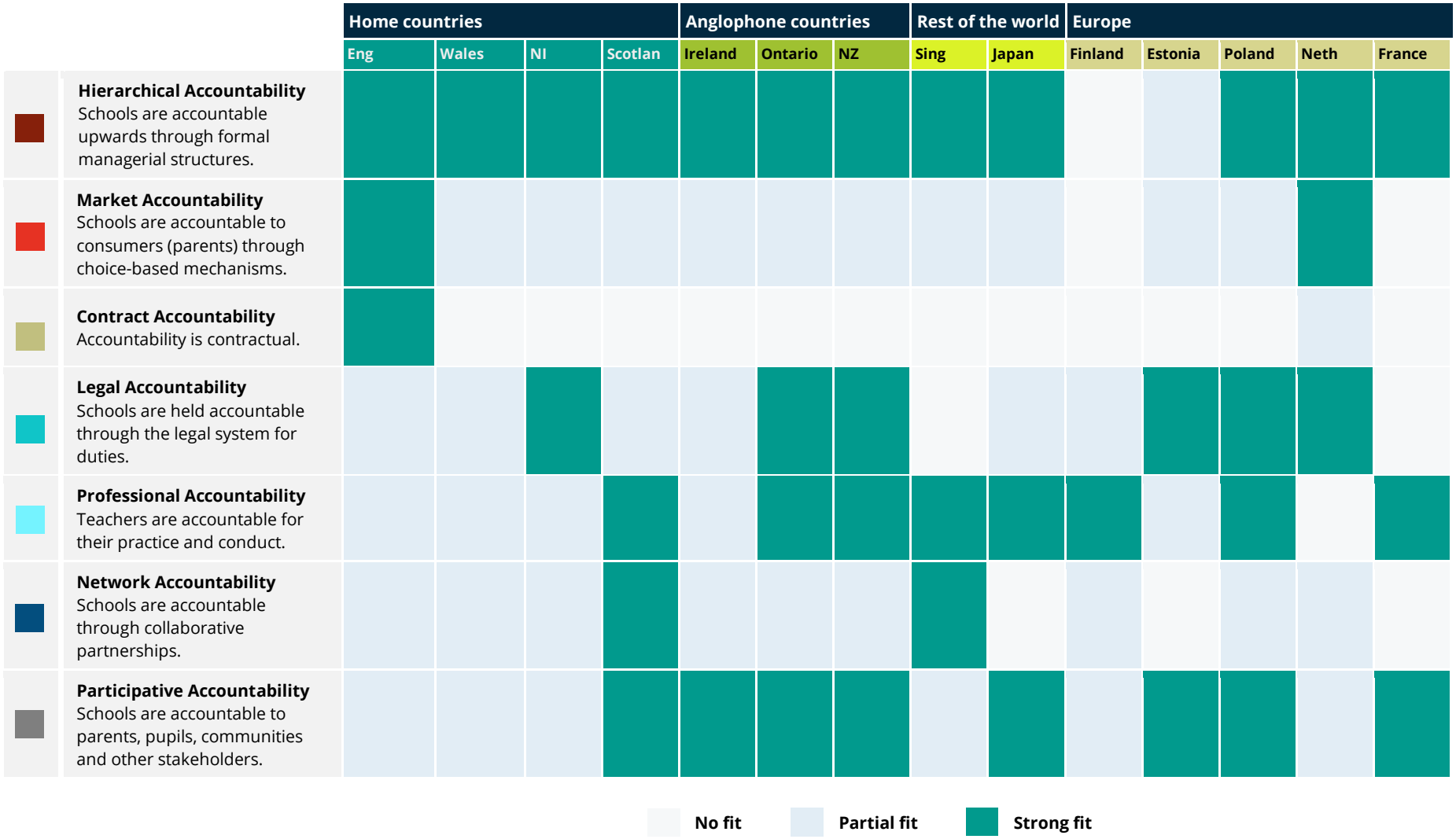
Network Accountability:
Schools are accountable through collaborative partnerships with formally independent groups, schools or individuals. Accountability within these networks often involves dialogue, and sanctions are generally weak, relying on constructive dialogue. However, loss of stakeholder support or exclusion from the network by peers is possible.

As noted earlier, this approach is present in most countries through collaborative networks, but these are often informal, localised and professional development focused. Networks can involve other actors from other services (like Children's Centres).[50] Given the similarities between network accountability and horizontal professional accountability the two are grouped together in the 'common approaches' described below.

Participative Accountability:
Schools are accountable to parents, pupils, and communities through dialogue, governing bodies, school councils. There are few formal sanctions beyond possible loss of stakeholder support.

This is an important feature of around half of the reviewed systems. The recent expansion of the 'community school' programme in Japan is an example of deepening this approach, whereas requirements for stakeholder input in governance have been scaled back in England. Although 'participatory accountability' does not normally include electoral or political accountability (for example at a ministry or local government level), Ontario's directly elected school boards can reasonably be included within this category.

[50] This form of accountability has been described as 'polycentric' elsewhere (Brown et al. 2020; Ehren et al. 2017) but use of the term in the current study aligns with literature on educational governance (Ball, 2009; Hughson and Menzies, 2026)


Home countries
Anglophone countries
Rest of the world
Europe
Eng
Wales
NI
Scotlan
Ireland
Ontario
NZ
Sing
Japan
Finland
Estonia
Poland
Neth
France
Hierarchical Accountability
Schools are accountable upwards through formal managerial structures.
Market Accountability
Schools are accountable to consumers (parents) through choice-based mechanisms.
Contract Accountability
Accountability is contractual.
Legal Accountability
Schools are held accountable through the legal system for duties.
Professional Accountability
Teachers are accountable for their practice and conduct.
Network Accountability
Schools are accountable through collaborative partnerships.
Participative Accountability
Schools are accountable to parents, pupils, communities and other stakeholders.
No fit
Partial fit
Strong fit

Accountability Webs

Given their tendency to combine so many different types of accountability, along with multiple relationships, accountability systems are best conceived of as webs.[51] The accountability webs below map out the main participants (actors and forums) that can be found across CES' fourteen jurisdictions and show how they are connected through various 'types' of relations. Doing so reveals:

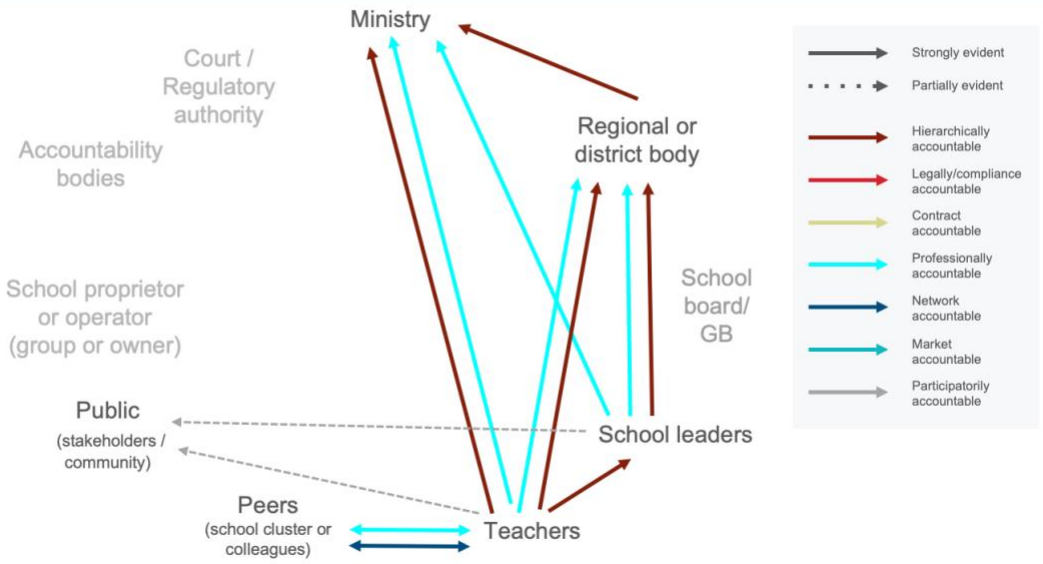
- Which actors and forums are involved in each jurisdiction, and the degree of centralisation, delegation or polycentrism.
- The prevalence of vertical, or horizontal, or diagonal relationships (as denoted by the direction of the arrows).
- The predominant relationship 'type' – including the sort of measure and likely consequences involved (as represented by the colour-coded arrows).

The webs therefore bring together multiple features of accountability systems and show how they are layered on top of each other.

While every effort has been made to ensure the accuracy of these webs, they are highly complex and classification judgements were often challenging. CES therefore particularly welcomes feedback on these webs from country experts and they will updated in the future.

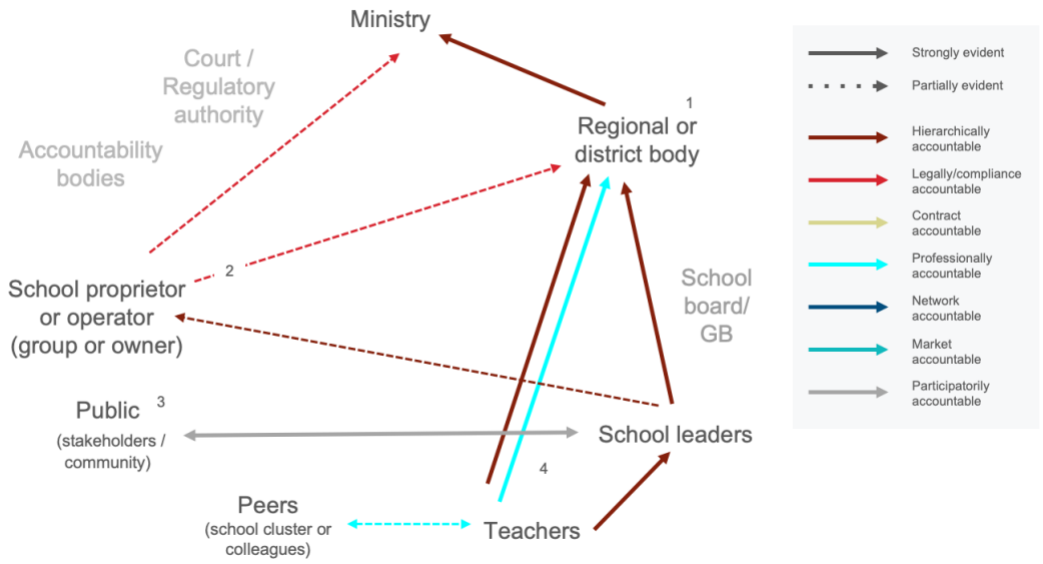
[51] Centre for Governance and Scrutiny. (2010).

Accountability web: Singapore



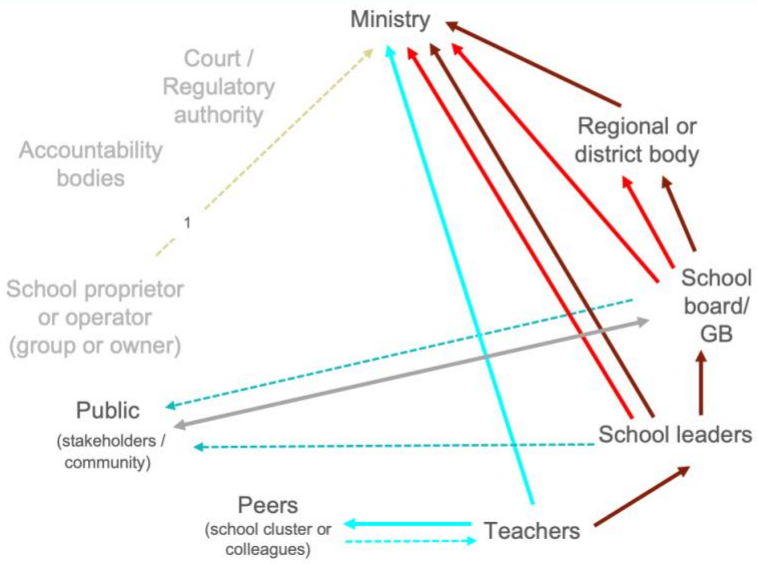
Accountability web: Japan

Notes
1. Prefecture and municipality + their boards
2. Private schools - particularly at upper secondary
3. Schools have a council but without strong accountability role.
4. Recertification requirements abolished in 2022. Teacher evaluation required by law. This is partly internal and managerial but some prefectural/ municipal involvement and they make job appointments
5. Famous for 'Lesson study' - but not necessarily accountability



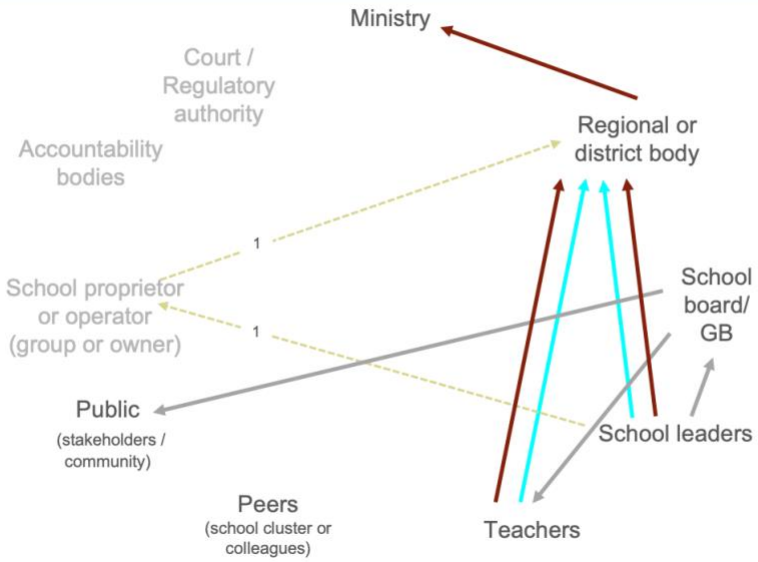
Accountability web: Estonia

Notes
1: Private schools: Estonia has a small number of privately owned schools that receive public funding and operate under a licence from the ministry.



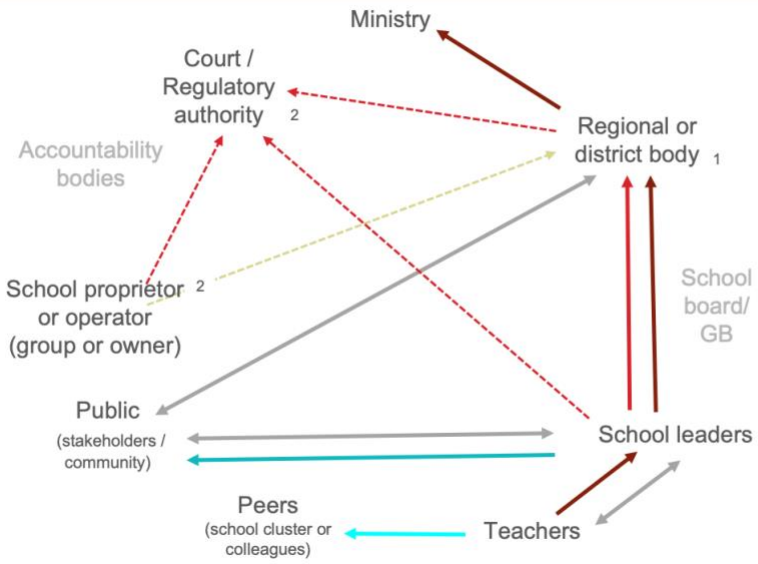
Accountability web: France

Notes
1: Schools must establish a contract agreement with the State and submit an annual progress report but policy expert, explains this is an administrative contract whereby the rectorat introduces its objectives - rather than an operating contract akin to a government funded independent institution.



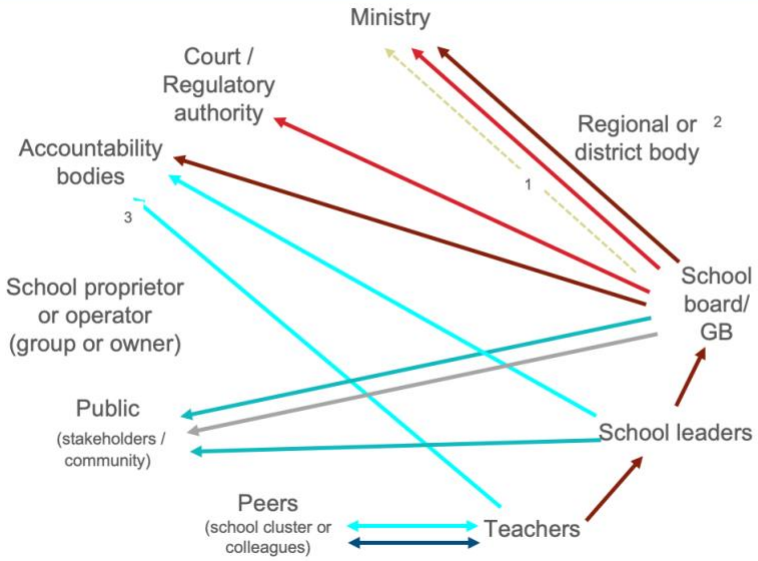
Accountability web: Poland

Notes
1 Schools are overseen both by local government and the REA. The REA is a direct appointment from the ministry
2. Non-public schools are legally required to meet various conditions including curriculum compliance. They can receive funding subject to meeting requirements so could be said to be 'contract' accountable.
3. Ombudsman and audit office



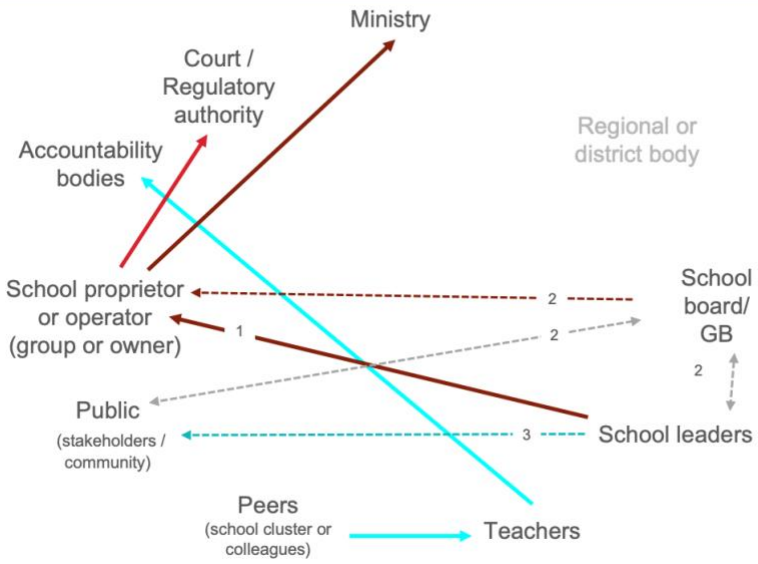
Accountability web: New Zealand

Notes
1: Charter schools
2. The ministry does have regional teams but it is not clear whether this constitutes an accountability related role.
3. The Teaching Council is classified as an accountability body but it combines this with a regulatory role.



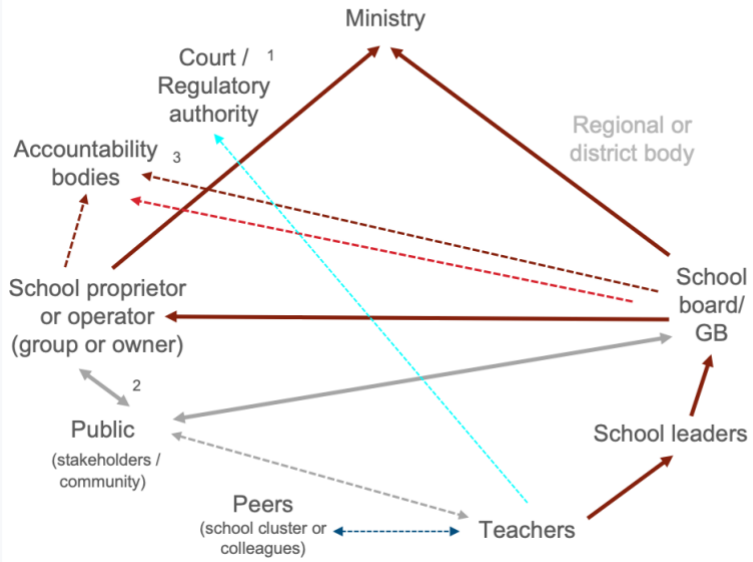
Accountability web: Ontario

Notes:
1: School boards are treated as school operators - see report.
2: Parent councils appear to have a primarily advisory role.
3. EQAO designs and administers tests that are published and may shape parental choices.



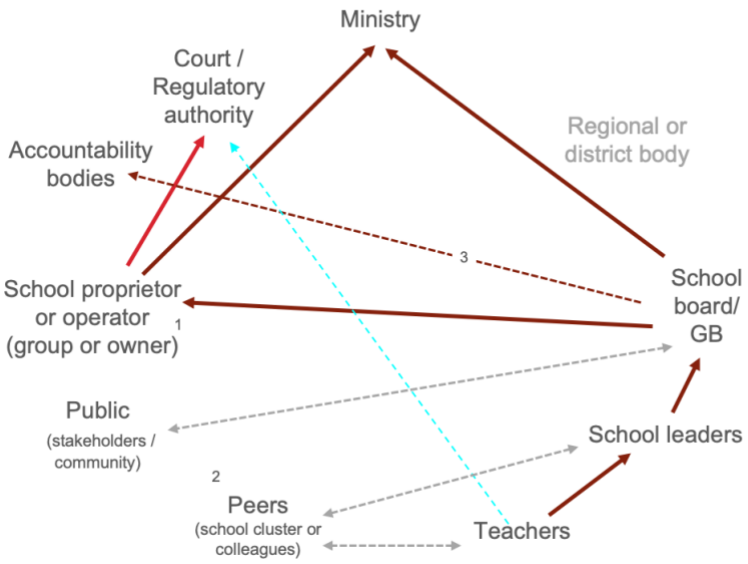
Accountability web: Ireland

Notes
1. The teaching Council primarily deals with malpractice. Statutory obligations overseen by bodies like "TUSLA" through reporting arrangements
2. Community involved in 'reconfiguration' to new patron - primarily focused on moving towards more multi denominational options rather than performance
3. Inspectors can make recommendations but the inspectorate is part of the ministry, and consequences come from recommendations to ministers.



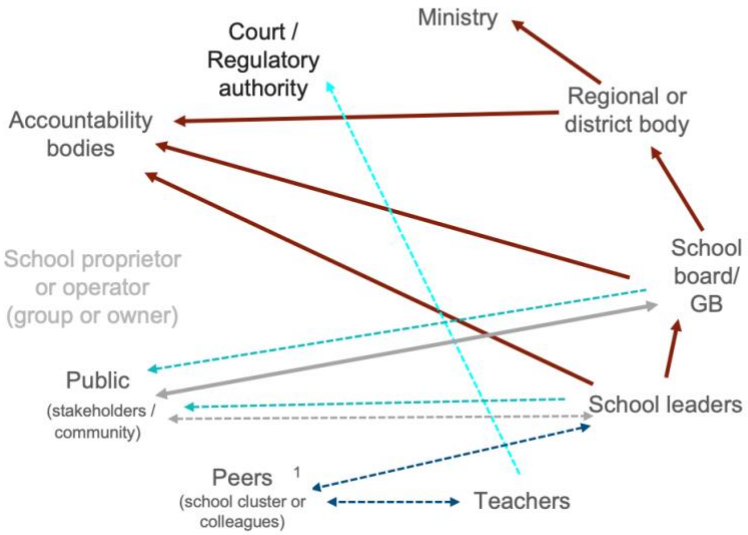
Accountability web: Northern Ireland

Notes
1. The EA and CCMS manage schools and are responsible for compliance.
2. Post primary schools are part of Area Learning Communities but these do not have an an explicit accountability role
3. Inspections are conducted by the ETI which is part of the ministry. The inspectorate makes judgements and consequences generally flow from Departmental decisions (via the EA in case of support)



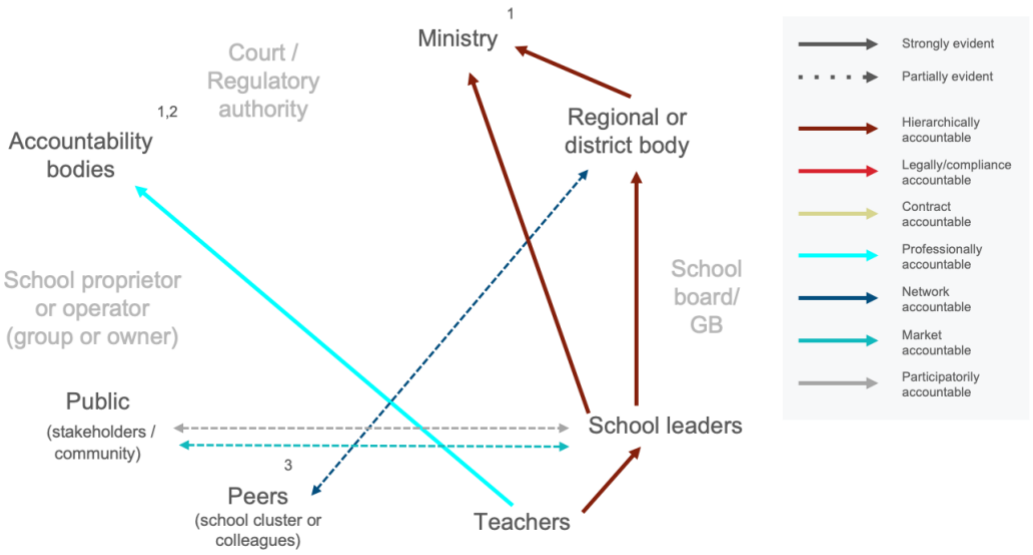
Accountability web: Wales

Notes
1. There has been a rebalancing of responsibilities, from regional consortia networks, back towards local authorities



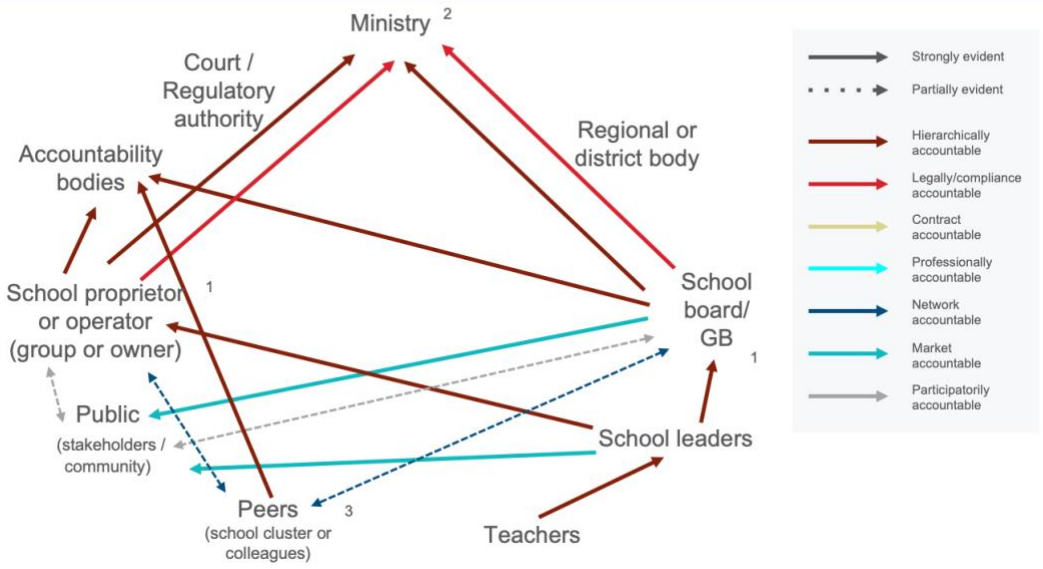
Accountability web: Scotland

Notes
1. Scotland's institutional structure is being reformed and will result in a more independent inspectorate and qualification body. Web based on previous system,
2. The GTCS describes itself as a 'regulator' but is classified as an accountability body here as its responsibilities extend beyond overseeing malpractice/basic standards.
3. Las work together in Regional Improvement Consortia
4. The degree to which Parent Councils play a role in accountability beyond consultation is unclear



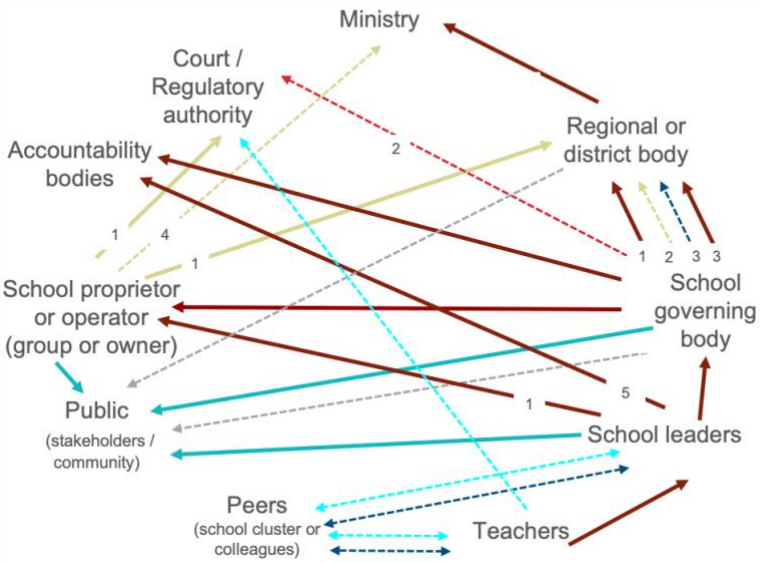
Accountability web: Netherlands

Notes
1: School boards can be for one or more schools. Boards representing multiple schools are categorised as school operators here.
2. School boards must report on compliance (legal) and can face loss of funding (contract). Reports go to DUO (Education Executive Agency) which this is part of the ministry.
3. Networks for inclusive education



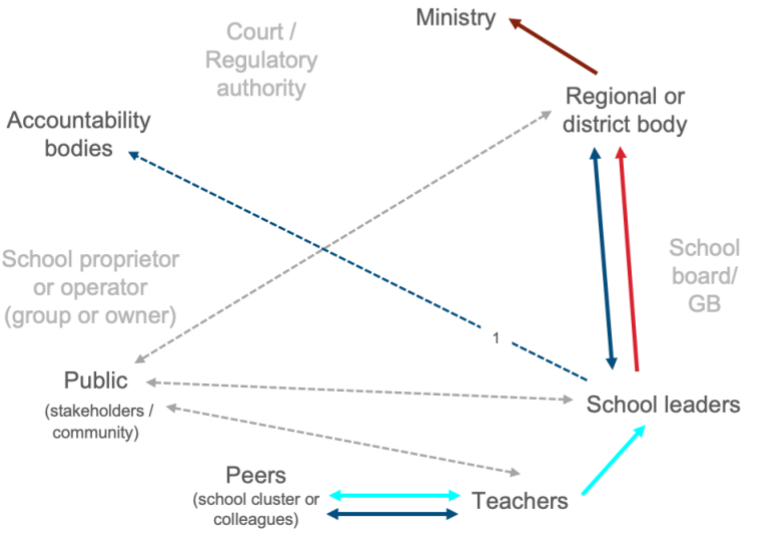
Accountability web: England

Notes
1. MAT schools *accountable to RSC / regulator / ministry*
2. SAT schools *accountable directly to RSC / regulator*
3. Maintained schools *accountable to LA and LA can be called to account by ministry*
4. Large MATs are overseen by the National rather than Regional Schools Commissioner .
5. Technically school leaders are held to account via their Trust or GB and one of the main inspection judgements relates to leadership and management.- and consequences are linked to this judgment.



Accountability web: Finland

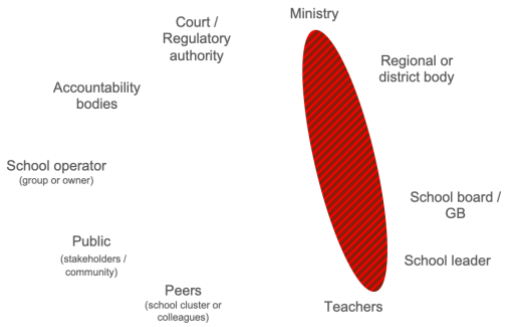
Notes
1. A sample of schools is evaluated (building on self evaluation), and data is reviewed by FINEEC (an independent unit within the Finnish National Agency for Education). This leads to formative feedback.Much of this activity relates to 'system-level' rather than 'school-level' accountability.
2. Although parents have numerous participative rights, they are said to rarely exercise these (SABER 2012)



Common approaches

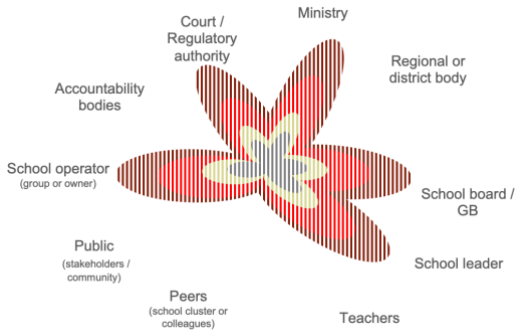
Reviewing the fourteen accountability webs reveals six common approaches. This refined typology broadly aligns with the accountability 'types' described above, but with adjustments to draw out some of the distinctive features identified.

Approach 1: Vertical chain of command Actors are accountable through a step-by-step management chain – for example from Head, to school board, to regional authority, to national government. This normally involves West's hierarchical accountability, but contractual and legal/compliance based accountability can also play a role.

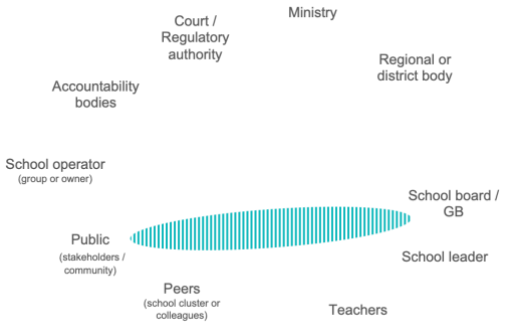


Approach 2: Vertical teacher accountability Teachers are directly accountable to regulatory authorities, accountability bodies, or the ministry. The focus can be either on compliance with teaching standards, or on development/professional practice.

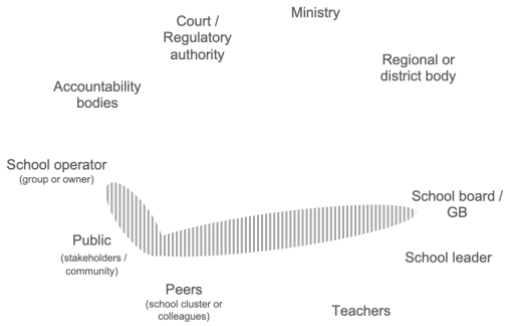
Approach 3: Polycentric accountability Multiple forums and actors are involved, and legal and contractual forms of accountability often play an important role alongside hierarchical relations. This approach leans towards governance and regulation rather than more hands-on management.



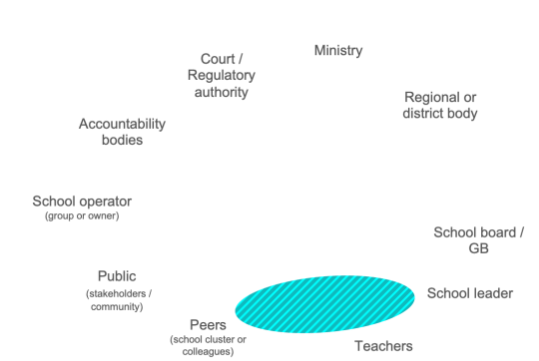
Approach 4: Market accountability The public holds school leaders or governing bodies to account through market/choice-based approaches.

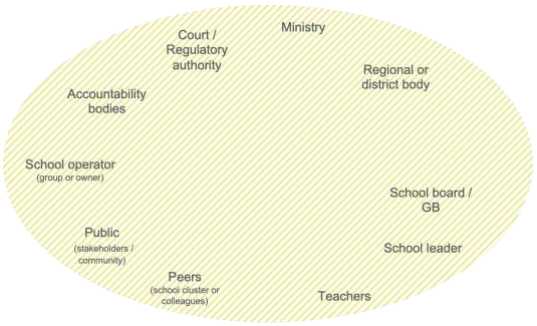


Approach 5: Participatory accountability to stakeholders
The public and community, most typically parents, participate directly in holding actors to account.



Approach 6: Horizontal professional accountability (internal and network based)
Teachers are held directly accountable for their professional practice by their peers, either internally or externally.



Approach 7: Collective system accountability The system as a whole is the actor as well as the forum. Individuals and public bodies are therefore held to account collectively in order to maximise performance at a jurisdiction level. This approach does not fall under narrower definitions of accountability.	

Part 2 References

NB: References for policy documents and additional sources can be found in the full and underlying reports and tables in the CES library on the CES website.

Ball, S. J. (2009). Academies in context: Politics, business and philanthropy and heterarchical governance. *Management in Education*, *23*(3), 100–103.

Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework1. *European Law Journal*, *13*(4), 447–468. https://doi.org/10.1111/j.1468-0386.2007.00378.x

Brown, M., McNamara, G., O'Hara, J., & Shevlin, P. (2020). Polycentric inspection: A catalyst for sparking positive interactions in educational networks. *Evaluation*, *26*(1), 76–97.
Centre for Governance and Scrutiny. (2010). *Accountability works! How we can make it better*. Centre for Governance and Scrutiny. https://www.cfgs.org.uk/wp-content/uploads/Accountability-works.pdf

De Grauwe, A. (2007). *Module 7. Alternative models in reforming school supervision*. UNESCO International Institute for Educational Planning. http://www.iiep.unesco.org/fileadmin/user_upload/Cap_Dev_Training/Training_Materials/Supervision/SUP_Mod7.pdf

Direction de l'évaluation, de la prospective et de la performance. (2025, September). *Indicateurs de valeur ajoutée des lycées*. Direction de l'évaluation, de La Prospective et de La Performance (DEPP). https://catalogue.depp.education.fr/index.php/catalog/274

Dubnick, M. (2005). Accountability and the promise of performance: In search of the mechanisms. *Public Performance & Management Review*, *28*(3), 376–417.

Ehren, M. (2025). *Context, mechanisms and outcomes of accountability systems in fourteen education systems. Working paper for the Centre for Education Systems*.

Ehren, M. C. M., Janssens, F. J. G., Brown, M., McNamara, G., O'Hara, J., & Shevlin, P. (2017). Evaluation and decentralised governance: Examples of inspections in polycentric education systems. *Journal of Educational Change*, *18*(3), 365–383.

European Commission. (2022). *Successful Pisa Stories in the EU*. https://doi.org/10.2766/429517

Hughson, T., & Menzies, L. (2026). Polycentrism in Education Policy Making in England and Aotearoa New Zealand: Strengths as well as Weaknesses? In C. Skerritt (Ed.), *Opportunities*

and challenges for policy advice and policy making in education: Future possibilities. Bloomsbury.

Inspectorate of Education (Netherlands). (2025). *Inspection Framework 2021 for the supervision of secondary education*. https://english.onderwijsinspectie.nl/documents/publications/2025/09/04/inspection-framework-secondary-education-2021

Maroy, C., & Voisin, A. (2017a). *UNESCO: think piece on accountability: background paper prepared for the 2017/8 Global education monitoring report.* https://doi.org/10.13140/RG.2.2.17900.18569

Maroy, C., & Voisin, A. (2017b). *UNESCO: think piece on accountability; Background paper prepared for the 2017/8 Global education monitoring report, Accountability in education: meeting our commitments; 2017*. https://doi.org/10.13140/RG.2.2.17900.18569

Mulgan, R. (2000). 'Accountability': An Ever-Expanding Concept? *Public Administration*, *78*(3), 555–573. https://doi.org/10.1111/1467-9299.00218

OECD. (2018). *TALIS 2018 Results (Volume II) Annex C Table II.3.30*.

Ofsted. (2023). *Education inspection framework for September 2023*. GOV.UK. https://www.gov.uk/government/publications/education-inspection-framework/education-inspection-framework-for-september-2023

O'Mara-Eves, A., Simon, A., Anand, K., Stansfield, C., Vigurs, C., & Ko, S. Y. (Jessica). (2025b). *Technical Report for Accountability Policy in Secondary Education. EPPI Centre.*

Ontario Ministry of Education. (2013). *Policy/Program Memorandum No. 119: Developing and implementing equity and inclusive education policies in Ontario schools*. Government of Ontario. https://www.ontario.ca/document/education-ontario-policy-and-program-direction/policyprogram-memorandum-119

Poisson, M., & Jorgoni, E. (2025). *OPEN SCHOOL DATA IN EUROPEAN EDUCATION SYSTEMS Council of Europe Platform on Ethics, Transparency and Integrity in Education (ETINED)-Volume 9 Council of Europe*. http://book.coe.int

SICI. (2024). *Country Profile Dutch Inspectorate of Education*. www.sici-inspectorates.eu/swfiles/files/SICI_-_Country_Profile_Template_the-Netherlands_27022024-1.pdf

Sinclair, A. (1995). The chameleon of accountability: Forms and discourses. *Accounting, Organizations and Society*, *20*(2–3), 219–237.

Udahemuka, M. (2016). *Signal loss: What we know about school performance*. New Zealand Initiative.

West, A., Mattei, P., & Roberts, J. (2011). Accountability and Sanctions in English Schools. *British Journal of Educational Studies*, *59*(1), 41–62. https://doi.org/10.1080/00071005.2010.529416

Part 3: The development and evaluation of accountability policy

Findings in the following sections draw on a policy review by the EPPI centre.[52]

RQ3: Policy design

Policy development processes are rarely well reported within public documentation. Findings in this section are therefore tentative and we have mainly focused on overarching trends and divergences. A full treatment would ideally involve primary research with policy makers. Inspection policy is normally one of the better documented areas of accountability policy so it dominates the analysis in this section.

Most countries follow a similar process – at least superficially – when developing their accountability policies.

- The ministry initiates major changes and sets the overall direction.
- Changes are then developed in partnership with accountability authorities such as the inspectorate (where these are independent), or with the relevant ministerial divisions in more centralised systems.
- Technical details and guidance – for example the contents of a new inspection framework – are then developed by accountability authorities or specialists with approval from the ministry**.** Larger changes may be scrutinised or ratified by parliament.

The extent to which policy development is ministry-led or accountability authority-led seems to vary:

- In some jurisdictions, policy documentation suggests the process is ministry-led (as is the case in Japan and Singapore).
- In other countries (like England and Finland), accountability bodies appear to have greater independence, but this is combined with significant ministry oversight, guidance and/or sign-off.
- In France and the Netherlands, authorities appear to operate somewhat more independently. For example, in the Netherlands the ministry delegates much of its power to the inspectorate. In Scotland, ministers set strategic objectives and then (until recently) held the chief executive of Education Scotland accountable for delivering these – however Education Scotland is technically part of the ministry.

[52] O'Mara-Eves et al. (2025b)

The degree of centralisation in policy design may not always correspond to the degree of centralisation of policy delivery.

- In France, although inspection is carried out by the ministry and its inspectors, the Council for School Evaluation (Conseil d'Evaluation de l'Ecole – CEE) plays an important role in shaping accountability frameworks.

Parliament typically plays a role in overseeing accountability, but in some cases inspectorates are also directly accountable to the legislature. Parliaments typically sign off legislation that sets out statutory arrangements for accountability.

- In Ireland and England, the inspectorate is accountable to parliament.

Based on the reviewed documentation, the following four categories can tentatively be suggested, though the reality is more of a continuum. Drawing firm lines between different jurisdictions' approaches is difficult, since influence will often be hidden, and because different authorities (e.g. teacher certification bodies and inspectorates) have differing levels of autonomy.

Group 1: Ministry dominated	Group 2: Shared ministry- accountability body	Group 3: Significant independence or discretion for accountability body	Group 4: Local delegation under strategic guidance
Singapore Wales Northern Ireland Ireland Poland	England Wales Scotland Finland New Zealand Estonia	Netherlands France	Ontario Japan

In systems with local delegation, central government still sets strategic direction.

- In Ontario, the development of overarching obligations (or 'legislative intent') is centralised but school boards have considerable discretion over school evaluation.
- In Japan, central government now sets guidelines for school evaluation but boards (municipal or prefectoral) retain considerable autonomy.

There have been signs of centralisation of accountability in some jurisdictions, but a loosening of control elsewhere.

- In Ontario the 'Better Schools and Student Outcomes Act', (2023) strengthened the provincial government's authority over school boards.
- In France, the 'Loi pour une école de la confiance' sought to lend greater coherence to school evaluation, partly through the CEE.
- In New Zealand, the Education and Training Act 2020 introduced new requirements on school boards.

In contrast:

- In Ireland the government has been trying to heighten schools' autonomy and freedom by allowing them to set their own priorities in self-evaluations.
- In Wales, there has been some rebalancing of school improvement responsibilities from Regional Consortia towards Local Authorities.[53]
- Scotland is currently in the midst of reforms prompted by an OECD review, and the Muir review, which will result in greater independence for the inspectorate and exams regulator. This can be seen as a move away from a relatively state-centric system towards a more polycentric arrangement.
- A trend towards decentralisation has also been described in the Netherlands.

Spotlight the French Council for School Evaluation (Conseil d'Evaluation de l'école – CEE)

The French Council for School Evaluation is the central independent body responsible for coordinating and shaping school inspection in France. Established in 2020 under the 2019 Law for a School of Trust (Loi pour une école de la confiance), it designs the national Secondary School Evaluation Framework and produces reports, guidance, and recommendations.

Structurally, the council is composed of fourteen members, including a President, four parliamentary representatives, and six education experts from UNESCO, OECD, and various universities. Also included are three senior officers from the Ministry of Education: the Head of Education (DGESCO), the Head of the General Inspectorate of Education, Sport and Research (IGÉSR), and the Head of the Directorate of Evaluation, Forecasting and Performance Monitoring (DEPP).

The CSE develops the tools and methodologies underpinning school evaluation including Self-Evaluation Guides, Domain Toolboxes, and questionnaires for stakeholders. It also works with the DEPP to design data indicators characterising school profiles.
Alongside its responsibility for ensuring consistency and co-ordinating the national programme of evaluation, it analyses students' progress and measures of school equity, producing yearly reports on the national evaluation campaign, as part of its system-level oversight role.

Key triggers for reform include:

1. New education strategies and reviews. These are the most common triggers for reform, although other factors will often have contributed to the decision to conduct a review or launch a new strategy.

- OECD reviews are a particularly widespread influence over reform.

[53] Department for Education – Ireland (2022), Senedd Research (2025)

2. Concerns about curriculum narrowing and pressure or the burden associated with accountability.

- Concern about accountability pressure and curriculum narrowing helped trigger reforms in Wales, Singapore, England and Ontario.
- Concerns about the burden associated with accountability were also important in New Zealand, England and the Netherlands

3. 'Pisa shock' when countries are concerned about falling or persistently low performance in international surveys.

- Pisa 'shock' was a potential driver behind reforms in France, Wales and Northern Ireland. We return to the influence of the OECD and Pisa in 'Contexts and Effects'

4. A desire to promote consistency or alignment with other policy priorities

- Changes to accountability in Finland followed from reforms to the curriculum.

Think tanks and pressure groups are occasionally cited as influences on policy development. Examples include:

- Praxis in Estonia,
- People for Education in Ontario,
- The Education Policy Institute in England.
- Research and guidance from international bodies beyond the OECD, for example UNICEF in Singapore, and the EU Commission in several European jurisdictions

Approaches to reform may vary more *within* a country than *between* countries. Smaller, incremental changes and technical adjustments often emanate from an accountability body (or the relevant division within a ministry), whereas more significant 'punctuation' events[54] can require major legislation or political leadership:

- Despite the CEE's considerable autonomy in France, it was itself only established in 2020 as part of major government initiated legislation.[55]
- As an Executive Agency, Education Scotland has had considerable independence in shaping accountability policy. However, legislative changes in 2025 led to a major restructure, including the creation of an independent inspectorate, and the replacement of the Scottish Qualifications Authority (SQA) by Qualifications Scotland.
- Significant shifts in Singapore tend to follow from strategic initiatives that emanate from a ministerial – or even prime-ministerial – level, but the details of these changes are worked out in more incremental and bureaucratic ways.

[54] Baumgartner, F. R., Jones, B. D., & Mortensen, P. B. (2023). Punctuated equilibrium theory: Explaining stability and change in public policymaking. *Theories of the Policy Process*, 65–99.
[55] The so-called 'Law for a School of Trust (Loi pour une école de la confiance)'

RQ4: Policy review and evaluation

Accountability authorities, like inspectorates and exam regulators, are commonly expected to produce annual reports on their activity. Incremental adjustments are normally made in response to these reviews and **t**hematic reports produced by accountability bodies can also contribute to policy adjustments.

Most countries have advisory groups or oversight bodies that feed into evaluations, reviews and reforms of accountability policy. However, the degree to which these bodies are formalised or permanent varies. Examples include:

- **Accountability bodies' governing boards:** For example, the board of the Education Quality and Accountability Office (EQAO) in Ontario, and Ofsted's board in England.
- **Boards or councils that advise on education as a whole (or with a relatively broad remit):** These can input into accountability policy. They include Japan's Central Council for Education; HAKA in Estonia; Ireland's National Council for Curriculum and Assessment; and Education Scotland's advisory board as well as Scotland's International Council of Education Advisors (which publishes an annual report).
- **Institutions that oversee or shape accountability policy specifically**: Notably FINEEC in Finland and the CEE in France.
- **Other government institutions such as Ombudsmen, Commissions, and auditors:** Ontario's Ombudsman was instrumental in the creation of Integrity Commissioners for school boards. In New Zealand the State Services Commission played a role in ERO's external review.
- **Semi-permanent bodies convened to oversee a period of reform:** These include Wales' Change Board, and Northern Ireland's Independent Review of Education.
- **Independent institutions with a long-term role:** The ministry in Singapore has a long-running close relationship with the National Institute of Education which continuously feeds emerging evidence and expertise into policy development.

The following table lists examples of somewhat established bodies, but is not exhaustive.

England	Ofsted board
Wales	Change board with advisory group oversaw reforms from 2019 (no longer in existence)
Northern Ireland	Independent review of education Panel launched in 2021
Scotland	Education Scotland advisory board International Council of Education Advisers (ICEA) – established in 2016 Publishes an annual report Scottish Education Council
Ireland	National Council for Curriculum and Assessment
Ontario	Education Quality and Accountability Office Board Ombudsman (eg led to creation of Integrity Commissioners) State Services commission involved in ERO external review
New Zealand	State Services commission involved in ERO external review
Singapore	Close working with National Institute of Education
Japan	Central Council for Education
Finland	FINEEC leadership and council members appointed by government for four-year terms Audit office
Estonia	Estonian Quality Agency for Education (HAKA)
Poland	–
The Netherlands	Education Council
France	Conseil d'évaluation de l'école

Alongside these relatively established groups, countries also frequently convene more adhoc groups that contribute to reform.

- Japan has a relatively established process for appointing expert review committees.
- Estonia has convened expert groups to create vision papers.
- Seven external reference groups were set up as part of Ofsted's recent reforms to inspection in England; and a Commission on Assessment Without Levels was established in 2015.

All countries consult widely as they develop their policy, but it is impossible to gauge how meaningful, deep or impactful this is from policy documentation.

Consultation normally involves unions, teachers and parents. Pupils and academics are also frequently consulted.

- Academics have considerable input into accountability policy making in Singapore (as with other areas of education policy) thanks to the aforementioned links between the National Institute of Education (NIE) and the ministry. As noted above, France's CEE is also structured to draw in academic insight. Academic involvement in accountability policy also appears to be high in Estonia.
- Northern Ireland is unusual in having established a dedicated pupil reference group.

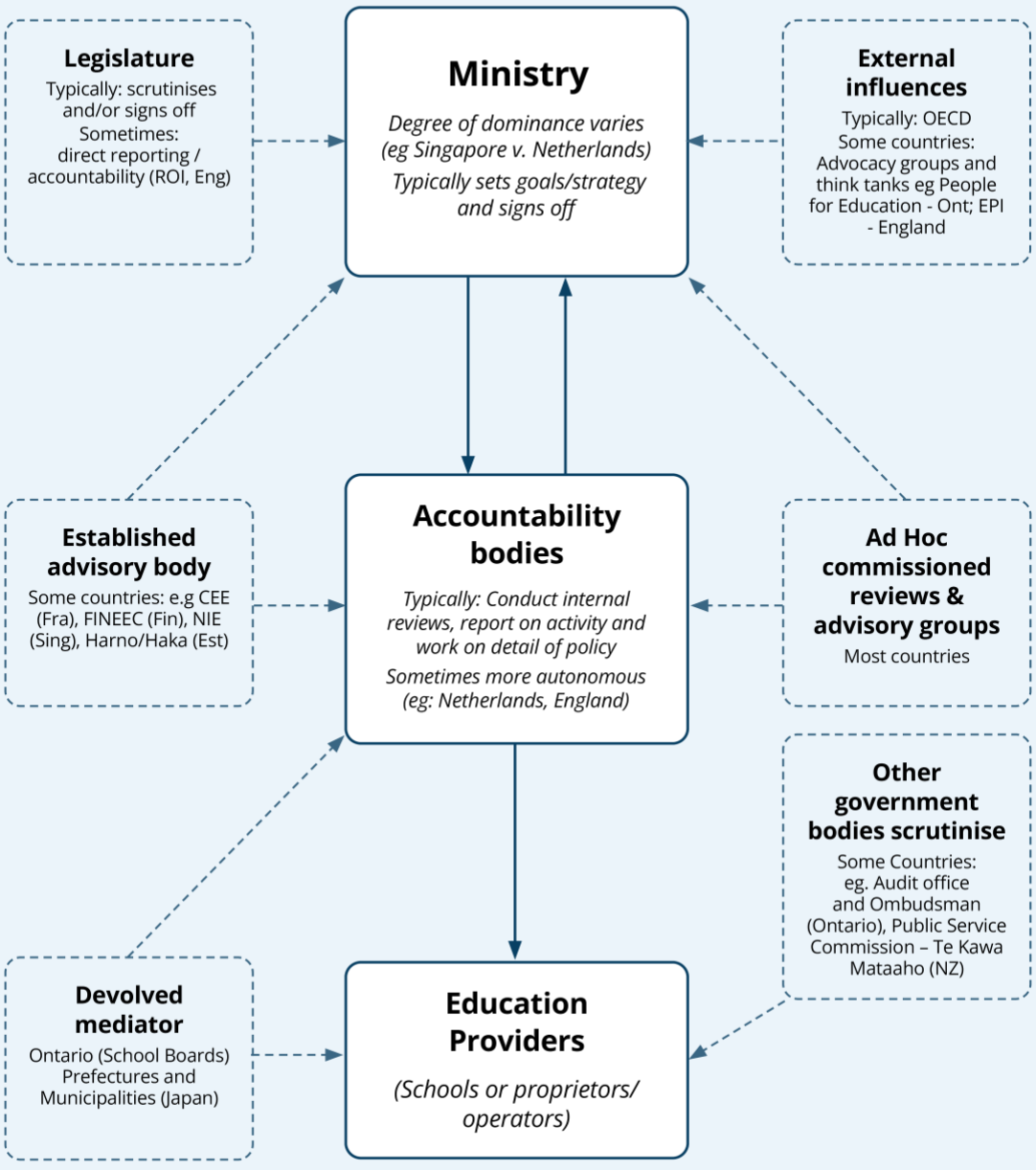
- In keeping with their cultural and demographic context, Ontario and New Zealand consult with indigenous groups.

In some cases reviews are ad hoc, but some countries have established mechanisms for conducting reviews on a more regular basis.

- In Finland, there is a planned timetable for reviewing accountability policies through the National Education Evaluation Plan, which operates on a four-year cycle.
- In Singapore, reforms are often linked to national 'Masterplans' which are reviewed every few years.
- Other jurisdictions, like England, run scheduled reviews, but do not follow a regular, planned cycle.

Policy is normally refined in response to pilots. These tend to be run whenever major reforms are introduced.

Designing and reviewing accountability policy



Part 3 References

NB: References for policy documents and additional sources can be found in the full and underlying reports and tables in the CES library on the CES website.

Baumgartner, F. R., Jones, B. D., & Mortensen, P. B. (2023). Punctuated equilibrium theory: Explaining stability and change in public policymaking. *Theories of the Policy Process*, 65–99.

Department for Education (Ireland). (2022). *School Self-Evaluation: Next Steps* . https://www.pdst.ie/sites/default/files/SSE_next_steps.pd

O'Mara-Eves, A., Simon, A., Anand, K., Stansfield, C., Vigurs, C., & Ko, S. Y. (Jessica). (2025b). *Technical Report for Accountability Policy in Secondary Education. EPPI Centre.*

Senedd Research. (2025). *What's happening to school improvement services?* https://research.senedd.wales/research-articles/what-s-happening-to-school-improvement-services

Part 4: The contexts for, and effects of accountability policy

This part of the report draws together findings from three different reviews of evidence on accountability:

- A thematic synthesis
- A realist synthesis
- A focused review of experimental and quasi-experimental studies

RQ5: Accountability contexts

Accountability systems are profoundly influenced by jurisdictions' historic, cultural, political and economic context. This section summarises the main findings from a review of studies that document these influences, conducted by the EPPI centre – within which full references can be found.[56]

History

History influences governance patterns and the role of the state, as well as ongoing debates about accountability.

- Both Estonia and Poland's reform trajectories have been interpreted as multi-phased reactions to their Soviet/Communist pasts.
    - Post-1990s reforms in Poland sought to dismantle Soviet-style centralisation by introducing school autonomy, parent choice, and decentralised accountability. Responsibilities for oversight moved from central authorities to local governments, and performance-based evaluations rose to prominence. However, mistrust of evaluation and initial resistance to accountability mechanisms persisted, perhaps reflecting lingering fears from the authoritarian era and resulting interpretations of accountability as forms of control (and perhaps oppression). A new phase is underway, which seeks to balance state oversight with local control, through standardized exams, collaborative self-evaluation, and periodic national inspections.
    - Estonia transitioned from Soviet-style centralised control, to a decentralised structure. This set the stage for the current emphasis on professional trust, school autonomy, and self-evaluation – with the central government setting broad policy frameworks but avoiding routine external inspections.

- The Meiji Restoration is argued to have launched Japan's emphasis on state-guided, centralised education as a nation-building tool. This was followed by reforms in the

[56] O'Mara-Eves, A. et al (2025)

1990s which introduced statutory duties for evaluation, public reporting, and municipal governance reorganization – echoing global trends towards 'New Public Management'. New Zealand also underwent similar reforms from the late 1980s onwards which were ushered in by 'Tomorrow's Schools'. This was a set of policies that introduced autonomous self-managing schools with oversight from elected boards, combined with data-driven accountability and market competition. However criticisms of historical and continuing inequities rooted in colonial history – particularly from indigenous Māori perspectives – have challenged the legitimacy and effectiveness of these systems.

Each of these cases demonstrates the tendency for different paradigms to be layered on top of each other over time. As a result, most systems blend 'old and new logics' – for example , market mechanisms combined with traditional bureaucratic approaches. This leads to what is known as 'accountability policy hybridity', and potentially, to contradiction – a theme we shall return to.

Politics and Economics

The degree of political centralisation within a jurisdiction often reflects historic path dependency.[57]

- Previous studies suggest that Finland, Estonia, and New Zealand all share an emphasis on school or municipal autonomy, paired with professional trust and self-evaluation.
- Other countries embody more centralised approaches to governance, with governments or municipalities playing a hands-on role in decisions like teacher appointments. This approach has persisted in France despite attempts at decentralisation.
- Systems like England and Singapore (particularly under its previous EPMS system) make greater use of external control through external mandates, national targets, and performance-based accountability. However, high levels of autonomy – in terms of formal school based management – are often paired with high levels of control through accountability (as noted in this report's Overarching Findings), either to ensure alignment with central policy goals, or as a means of checking whether schools are performing well (or a combination of the two). School autonomy is constitutionally enshrined in the Netherlands and as the OECD explains, "the system is underpinned by: a high level of decentralisation… balanced by a solid accountability system that includes a national examination and a strong Inspectorate of Education; school financing

[57] Kay (2012)

that supports disadvantaged students; experimentation and innovation; and good data and research."

The belief that markets and competition could yield greater efficiency became increasingly dominant over the 1980s and 1990s.

- For example, in England there was a shift from post-war welfarism to market-driven models in the late twentieth century. This precipitated moves towards performance data, league tables, competitive comparison and targets. Other systems (for example Finland) have resisted these approaches, or have hybridised market- and outcome-focused reforms with other approaches, as the Netherlands, France and Scotland have done.

Resource distribution, budget constraints, and funding mechanisms also shape the feasibility and character of accountability. Researchers suggest that resource inequities may exacerbate challenges to equitable accountability, particularly in systems impacted by austerity or historical crises, or in competitive systems where resources move towards better rated schools.

Socio-cultural influences

Whereas some countries are said to prioritise inclusion and consensus, others are said to be anchored in meritocracy and achievement.

- Scotland, Ontario and Finland are said to prioritise inclusion and consensus, for example through a long-running emphasis on educating students with Special Educational Needs within the mainstream (see CES SEND policy report).
- Finland and Estonia's accountability systems reflect higher levels of professional trust and educator agency, according to many studies. This has been linked to these jurisdictions' limited use of sanctions, and an approach to governance that focuses on consensus and the state as a source of capacity-building support. In particular, Finland's system relies on school self-evaluation, peer review, and inclusive decision-making involving teachers, parents, and local authorities. Routine punitive measures are uncommon, with policy focused on supporting professional growth and collective improvement.

Local leaders, traditions, and socio-cultural factors can all mediate policy impact.

- Studies of accountability in France point to considerable resistance at the school level, partly aided by the country's strong union voice.

International influences

Nearly all jurisdictions have been shaped by international forces, particularly OECD/PISA benchmarks.

- Jurisdictions like England and Singapore have not only adopted or imported international policy models, but served as sources from which other countries borrow or adapt accountability policies and educational inspection frameworks. Elements of England's inspection model, league tables, and technocratic accountability have been taken up or explored in East Asia, Ireland, and Scotland.
- Singapore exported its approaches to teacher development and assessment across the Asia-Pacific region.
- Jurisdictions like Finland have engaged with international models selectively while resisting moves towards competition or market-based ranking. Established performance in international 'league tables' may reinforce confidence in the existing approach.
- The Netherlands has drawn on international models, but sought to engage in consensus-driven adaptation.

RQ6: The effects of accountability systems

In this section we review research from across the CES jurisdictions, and beyond, to assess the evidence on educational accountability's effects.[58]

We explore effects in relation to:

- Student performance, in terms of attainment.[59]
- Wider pupil effects, including engagement, attendance, motivation to learn and longer term behavioural and employment outcomes.
- School and teacher practices.
- Workforce effects including teacher wellbeing and turnover.
- Pupil distribution.

Throughout, we consider overall and distributional effects (including their implications for equity). We particularly focus on the mechanisms that might drive effects, and how these are shaped by, or depend on, context and circumstances.

[58] See methods: The EPPI Centre thematic review and Melanie Ehren's realist review focused on the CES jurisdictions but Simon Burgess' review of quantitative studies drew from across the world, given the scarcity of research meeting the studies' criteria.
[59] The review sought evidence on wider measures of performance including so-called '21st century skills' such as creativity but studies tended to focus on related themes, including innovation in teaching methods and curriculum narrowing.

A recurring theme is that accountability's effects are heavily intertwined: system alignment, school competition and pressure to focus on examined content are well-documented consequences of certain forms of accountability. However these might also play a role in driving increases in school performance (as measured by test-scores), and this may be desired or undesired, depending on policy makers' intentions, and their chosen policy strategies.

Pupil attainment

Studies of accountability's impact on pupil attainment tend to relate to performance-based accountability systems in which test scores are linked to consequences (whether market-based or hierarchical). This is partly because by their nature, these systems have information about standardised student outcomes available to study. The vast majority of studies of this type were conducted in the US as part of 'No Child Left Behind' (NCLB). NCLB was a policy introduced in 2002 that linked federal funding to the adoption of test-based accountability. These studies tend to use large samples and be quantitative in nature.

Positive effects on pupil attainment were found in most of the reviewed studies of test-based accountability in the US.[60] Similar effects have also been found elsewhere: a study based on a natural experiment when policies in Wales and England diverged, found that pupil attainment in Wales fell when performance accountability was reduced.[61] In Brazil, publication of school-level results in the national exams at the end of secondary school was found to increase attainment, particularly in private schools[62]. Similar effects were found in Pakistan and Tanzania.[63]

Effects on attainment are not uniform, but the distribution of performance gains may be equity enhancing. However, these findings should be read in conjunction with findings on other effects that may be equity inhibiting. Numerous studies suggest that positive effects on attainment are concentrated within lower performing schools,[64] and that there may be no effect in higher performing (or higher 'rated') schools.[65] The aforementioned study of England and Wales also showed a greater positive impact on low-performing pupils than on their higher-achieving peers[66]. There is also heterogeneity between jurisdictions. One large-scale study tracked the introduction of accountability policies across multiple jurisdictions and found that positive effects were stronger in lower

[60] Hanushek and Raymond (2005); Chiang, 2009; Figlio & Rouse, 2006; Rockoff & Turner, 2010; Rouse et al., 2013; Dee, 2020; McElroy (2023); Deming et al., 2016; Eren et al (2023); Feng et al. (2018)
[61] Burgess et al. (2013)
[62] Camargo et al. (2018)
[63] Andrabi et al (2017), Cilliers et al (2021)
[64] Cilliers et al (2021); Deming et al., 2016
[65] Feng et al. (2018)
[66] Burgess et al. (2013)

and average-performing contexts, and that there may even be negative effects when performance based accountability is introduced in high-performing contexts.[67] This suggests there may potentially be a ceiling to effects. Relatedly, another study found that accountability policies (primarily those involving test-based accountability) were associated with improved pupil performance overall, but not in higher income jurisdictions.[68] On the other hand, one US study found that positive effects in schools at risk of a low rating were reversed among low-performing students.[69] Studies in Canada have also pointed out that widespread inequalities persist regardless of accountability policies, and that initially positive effects may fade and give way to more negative consequences.[70]

Effects on pupil attainment are sensitive to detailed features of the policy design, since these affect incentives and hence school behaviour. Key design considerations include the structure of the measure and how results are communicated:

- **Where performance measures focus on the number of students exceeding a threshold, schools may focus on pupils near that threshold, such that attainment gains are concentrated among these pupils.** One study demonstrated this by analysing shifts in league table metrics used in England. The authors found that schools respond precisely to accountability incentive structures since, once the incentive to focus on near-threshold students was removed, these 'borderline' pupils made relatively less progress, whereas other student groups benefitted significantly. Policy-makers therefore need to be careful to design measures that reflect their policy goals: if they want to increase the proportion of pupils meeting basic standards, then threshold measures might be appropriate; if they want to drive improvements across the board, then thresholds need to be avoided.[71]

    - **Performance gains in tested subjects appear to be greater when measures are based on standardised assessments; when measures are communicated to the public in a clear manner; and when there is more competition.** This evidence suggests that market-based competitive mechanisms may play a role in raising pupil attainment, as might pressure which drives schools to focus more on measured outcomes.
    - **Comparability and standardisation:** Improvements in pupil attainment were greater when accountability was based on more comparable and standardised measures according to comparative international research on

[67] Bergbauer et al (2024)
[68] Torres (2021)
[69] Deming et al (2016)
[70] Eizadirad, A. et al. (2016) Rezai-Rashti, G. M., & Segeren, A. (2023).
[71] Menzies et al (2023) p23

accountability systems.[72] Another US study found that when teacher evaluations were uncoupled from pupil outcomes attainment subsequently declined.[73]

- **School proximity and competition:** Competition seemed to drive pupil attainment gains in Japan, where a natural experiment showed that when test-based information was publicly available, student achievement increased more in areas where there were a larger number of schools nearby.[74]
- **Clear conspicuous reporting:** Performance effects were only found when school ratings were conspicuously publicly reported, according to a third US study. This was despite the fact that the less conspicuous system was coupled with a package of support. After following up their quantitative analysis with interviews, the authors concluded that clearer, more conspicuous metrics provided school leaders with "an obvious target for school improvement activities" and "a sense of urgency".[75] This suggests that at least accountability pressure was a more significant mechanism in driving performance gains than support. But of course, this may depend on the nature of the support.

The mechanisms driving these effects may relate to controversial secondary effects:

- Increased focus on accountability metrics may explain attainment gains, but 'narrowing' and 'pressure' are precisely what other systems seek to avoid.
- Pupils moving to higher performing schools may be what drives attainment gains in choice-based systems. This may be the intention when seeking to harness market forces, but it can also have inequitable effects by driving segregation and making it harder for struggling schools to improve (for example by making it harder to recruit teachers).[76]
- Teachers leaving and being replaced by more effective colleagues.
- Test-based accountability can incentivise the exclusion of lower-performing students and encourage teachers to cheat.

We return to these effects below.

Other, more horizontal forms of accountability may also have a positive impact on pupil attainment. One small-scale study in Northern Ireland (28 survey responses and 18 interviews) of an area-based collaboration in which inspection took place at a network level, highlighted improvements in pupil attainment. Although the authors note that

[72] Bergbauer et al (2024)
[73] Hanushek et al (2024)
[74] Morozumi and Tanaka, 2023
[75] Atchison et al (2025) p17, p1
[76] Ehren 2025

"it would be naive to suggest that successive yearly increases in examination results can be directly attributed to the creation of the polycentric network", they also report that interviewees "unambiguously" linked improvement in exam results to the network's work.[77] Unusually, this study focused on an initiative that combined hierarchical forms of accountability (through inspection) with horizontal professional approaches (through the area-based network).

Attainment may also improve when teachers are directly held to account for their practice – as is the case in Singapore. One small-scale study suggests that incentive and pressure-based mechanisms may play a role in driving these effects. However, some teachers may respond by being demotivated. As one teacher commented: "it's a double-edged sword. While it may raise the standard and it motivates some… it will also demoralise some".[78] Others suggest that this pressure may be as much derived from cultural norms as by the accountability system.[79]

Wider and longer term pupil effects

Studies have explored accountability's effects on pupils behaviour and attitudes; attendance; and longer term outcomes, including health-related indicators. Once again, the bulk of this research comes from large-scale quantitative studies in the US, linked to NCLB.

In certain contexts, accountability and the pressure associated with it might have a positive impact on attendance. In South Carolina, schools receiving a negative rating were found to subsequently improve retention of students (in other words, dropout rates were reduced). Over the border in North Carolina, another study concluded that accountability pressure may be passed to students such that this "encourages students to show up at school and to do so on time"[80]. These effects were greater in schools where the NCLB framework specifically incentivised reductions in absenteeism. Research in a very different context – Pakistan – also highlighted increases in school enrolment in response to the introduction of school report cards, suggesting that incentive-based mechanisms may transfer between contexts. However, the North Carolina study also found a negative impact on anti-social pupil behaviour, including sexual offences and actions leading to suspension. From an equity point of view, it is worth noting that these behaviours were particularly concentrated among minority and low-performing students. The authors suggest that the mechanism behind these effects is likely to be schools' tendency to devote "time and resources to improving incentivised behaviors at the expense of ignoring other behaviors", and potentially by "tightening reporting standards for offenses".[81]

[77] Brown et al. 2020 p91-92
[78] Hwa 2020, p235
[79] Lee 2024
[80] Holbein and Ladd (2017) p55
[81] Holbein and Ladd (2017) p66

One study suggests that test-based accountability under NCLB may have *reduced* pupils' test anxiety and, perhaps relatedly, that there was no negative impact on pupils' enjoyment of learning. The authors suggest that this may be linked to schools focusing more on test preparation and motivational techniques, which may have helped to "alleviate student anxiety rather than exacerbate it."[82] Similar, potentially counterintuitive findings like this have also been found elsewhere.[83] On the other hand, teachers in one study from Ontario report a perception that test-based accountability increased student anxiety.[84]

Pressure from performance based accountability can translate into a negative impact on pupils' health. Studies have found an increase in the calorie count at lunchtime on test days,[85] a reduction in school time dedicated to exercise and a subsequent increase in pupil obesity,[86] along with increased diagnosis of ADHD and medical prescription.[87] Once again it is important to note that all three of these findings come from studies of NCLB and that they particularly relate to schools facing the risk of accountability consequences.

There is some indication that accountability's effects may persist well after pupils finish school. Long run studies all emanate from the US and focus on NCLB related initiatives. Positive effects were found on measures of future education (e.g. college completion) and future income, but labour market effects were sometimes not statistically significant, and there was considerable divergence between measures, and between different groups of students[88]. One study also found reductions in welfare dependency and future criminal activity.[89]

[82] Reback et al (2014) p230
[83] For a brief review of the (highly contradictory) evidence on the relationship between testing and pupil wellbeing see Menzies 2023 p17–19
[84] Hargreaves 2020 pp.413–414
[85] Figlio and Winicki (2005)
[86] Anderson et al. (2017)
[87] Bokhari and Schneider (2011)
[88] Deming et al. 2016, McElroy (2023), Mansfield and Slichter (2021)
[89] Eren et al (2022)

Teacher and school practices

There is overwhelming evidence that accountability changes school and teacher practices, whether by influencing pedagogical approaches, or by incentivising so-called 'strategic behaviour'.[90] As noted earlier, these teacher or school behaviours can be seen as an effect of accountability, or as mechanisms driving pupil-level effects.

Alignment or compliance?

Schools and teachers' responses can be broadly grouped under two mechanisms: standardisation and isomorphism – whereby "accountability creates pressure, competition, and compliance" and organisational learning which is "driven by reflection, trust, and internal motivation for improvement". Which mechanism – and which associated effects dominate differs according to the form of accountability involved, and the implementation context. Thus, even within the same jurisdiction, schools "in deprived areas with a failing outcome who face local competition over student numbers" might experience more pressure, and therefore respond through standardisation and compliance; while schools performing highly on the accountability metric and serving a more affluent student population might respond through organisational learning.[91]

Many studies find that accountability reduces innovation in teaching (and imply that this is undesirable). This pattern tends to be linked to hierarchical accountability and the type of high-threat contexts that drive isomorphism – for example, in England, where teachers report that a sense of having to be 'ofsted ready' inhibits their teaching practice.[92] Dutch studies of risk-based inspections coupled with support illustrate the different ways that – what is ostensibly the same accountability approach – can lead to differing responses. Whereas some schools respond to these inspections defensively by prioritising compliance rather than innovation; others engage in reflection, evaluation and improvement.[93]

Schools' and teachers' responses to accountability appear to differ depending on various contextual factors. In effect, schools:

"experience different accountability regimes depending on contextual factors such as inspection results, staff capacity, and student composition. Schools in deprived areas with a failing outcome who face local competition over student numbers will for example experience more pressure

[90] Ehren (2025), Burgess (2025)
[91] Ehren (2025) p3
[92] Clapham, (2015)
[93] Burns, T., F. Köster and M. Fuster (2016), Education Governance in Action: Lessons from Case Studies, Educational Research and Innovation, OECD Publishing, Paris, https://doi.org/10.1787/9789264262829-en; Altrichter and Kemethofer, 2015; Jones et al. 2017

compared to their counterparts with a more affluent student population and good performance on the accountability metric."[94]

Accountability may be more likely to prompt organisational learning when:[95]

- There is an absence of pressure for immediate improvement.
- School or district inspectors support learning and improvement through professional conversations, "providing context-sensitive, unbiased, constructive and actionable feedback".
- Resources are allocated for improvement.
- It may also be more effective when it is organised in a way that aligns with "the values and professional capacity and autonomy of teachers, schools and the wider community".[96]

The discouraging of innovation is not limited to inspection-based accountability. Studies in Ontario have made similar claims in relation to the publication of pupil performance data and the resulting market-based pressure, combined with community-level stigmatisation of low-performing schools.[97]

One of accountability's most commonly documented impacts on teacher and school practice is 'teaching to the test', and a closely related tendency to 'narrow' the curriculum to focus more on assessed content. These effects have been found across numerous jurisdictions including the Netherlands[98], Estonia[99] and Ontario[100]. Numerous quantitative studies in the US have also demonstrated that schools respond to pressure from NCLB by increasing the amount of time dedicated to tested subjects and focusing more on topics and pupils that were likely to yield boosts in assessment results.[101] One small-scale cross-country survey uncovered a "clear association between increasing pressure in a school inspection system and an increase in the narrowing of the curriculum and instructional strategies in the school".[102]

There is also evidence of strategic behaviour and unhealthy competition linked to (vertical) teacher accountability.

[94] Ehren (2025) p3
[95] Ehren 2025, p19
[96] Ehren 2025, p23
[97] Hargreaves, (2020); Wang, (2017)
[98] Altrichter and Kemethofer, 2015; Ehren et al, 2015; Ehren et al, 2015b
[99] Sen and Maya (2022
[100] Jang and Sinclaire, (2018); Hargreaves, (2020)
[101] Reback et al. (2014) Figlio and Loeb (2011) Rouse et al. (2013)
[102] Jones et al. (2017) p818

- In Singapore – where teachers are held directly to account through the EPMS system – one teacher reported attempting to 'game' their rating by participating in activities the system would reward them for, like committee work.[103]
- The previous teacher evaluation system in Japan was said to have driven up competition and reduced collaboration between colleagues.[104]

Policy makers may not be displeased to learn that schools and teachers respond to the incentives created by accountability frameworks, if this helps align behaviour with policy goals. Pedagogically too, teaching to the test may not be such a bad thing if teachers are teaching to the right test – as Dylan Wiliam and Tim Oates have pointed out.[105]

'Isomorphic' and compliant responses can be mediated by school leaders resisting accountability pressures. Cross-country quantitative analysis of TALIS data has demonstrated principals' moderating role.[106] This tendency has been particularly noted in France, a context in which unions are particularly powerful; there is widespread resistance to performance-based accountability; and, government attempts to "reduce the opposition of teachers' unions and guarantee school principals room for manoeuvre" by institutionalising 'vagueness' in policy texts.[107]

Peer-to-peer, network-based and professional accountability models appear to drive different mechanisms, with contrasting effects. For example, small-scale studies in New Zealand and Belfast, Northern Ireland suggest that data informed professional enquiry (for example where teachers work together "to use data to examine and improve their own practice or overall school functioning"),[108] combined with support and direction from principals or inspectors, can encourage experimentation with new methods along with collaborative problem solving. On the other hand, an emphasis on outcome data-driven inquiry, can sometimes be experienced as a tension.[109] In Wales, a move towards more supportive and trust-based inspection involving peers as inspectors – along with less emphasis on grading – is viewed by some as having improved teachers' practice and wider knowledge of the system. It has also been said to give peer-inspectors an opportunity to develop their skills in areas like data analysis and lesson observation.[110] In England, peer inspectors were found to provide empathetic 'brokers' helping to explain judgements.[111]

[103] Hwa (2021)
[104] Jisu (2020)
[105] Quoted in Millard et al. (2017) p38
[106] Kim (2024)
[107] Moller et al (2024) Maroy (2021) p570
[108] Dyson 2020 (p3)
[109] Brown et al. 2020; Dyson, 2020
[110] Keane, (2023); Klimecki and Wilson, (2024)
[111] Moreton et al. 2017

Although policy makers may want to exercise a degree of control, they will generally want to avoid *hyper*-compliance. This is particularly important in avoiding more unambiguously negative effects, such as pupil exclusion and cheating, to which we now turn.

Schools can respond to accountability by removing potentially low-performing students from test year groups, or allocating pupils to specific streams, such as "special education" categories that are exempt from accountability measures. A considerable body of evidence for these effects comes from NCLB studies in the US[112] – and therefore relates to test-based accountability, but studies from Ontario,[113] Australia,[114] and Tanzania[115] have found similar effects. One US study focused on district level patterns and found that "negative ratings… led to the disenrollment of economically disadvantaged students… and a corresponding *(suspicious)* jump in test scores". The authors therefore underscore the potential role exclusion may have played in driving attainment gains.[116]

Evidence of cheating in tests has been uncovered in large-scale quantitative studies in Italy and Chicago. The Chicago study found that cheating was more widespread where accountability pressure was stronger.[117] In Ontario, documented "undesirable behaviours" include "opening the test before test day, providing students with questions in advance, erasing incorrect answers, teaching to the test,and circumventing the instructions for test administration".[118]

Workforce

Impact on the workforce is one of the most widely documented impacts of accountability. Effects fall into two main categories: those related to teacher and school leader well-being and workload (whether in terms of perceived pressure, or the resulting stress and anxiety); and those related to teacher turnover or mobility – which can sometimes be driven by the former.

Coercion and pressure from accountability can lead to anxiety, stress and frustration among teachers, and in some – but not all – circumstances this can undermine morale and confidence.[119] These effects have been found in multiple contexts – including the Netherlands, Northern Ireland, Singapore and England.

[112] Deming et al (2016) Eren et al (2022), Figlio and Getzler (2006) Gregg and Lavertu (2023)
[113] Rezai-Rashti and Segeren (2023) p269
[114] Coelli and Foster (2024)
[115] Cilliers et al (2021)
[116] Gregg and Lavertu (2023) p1. Emphasis added
[117] Jacob and Levitt (2003), Bertoni et al. (2013)
[118] Pinto (2016) p101,107
[119] Ehren 2025, P16; McClurg et al, 2025; Evans 2023

Professional accountability involving teacher evaluation can create workload and sometimes leads to compliant behaviour. One study of Ontario's Teacher Professional Appraisal system concluded that it "does not enhance teacher professionalism… [and] does not lead to self-actualization but rather causes trepidation and is time-consuming".[120] Two comparative studies of teacher accountability covering Singapore, Japan and Finland linked teacher evaluation programmes in the first two countries to compliant behaviour and gaming (though there may also be positive effects on pupil achievement in Singapore – as noted above). However, they found that in Finland, there appears to be limited impact on teachers' practice.[121] One of these studies concluded that teacher evaluation systems were more likely to build professional capacity when "coupled with school-level actions to support their professional growth and a school-level climate of shared responsibility".[122] It may be that such schemes have limited impact when teachers already feel a high degree of internal motivation – but there is also a risk that they might undermine intrinsic motivation.

Some approaches to inspection – particularly self-evaluation – may have more empowering effects.

- Ireland's approach to school inspection – drawing on structured self-evaluation combined with external oversight and reflection on school data – has been credited with empowering school leaders, and prompting reflection, planning, and action.[123] This may be linked to the country's otherwise "low-stakes accountability framework" – whereby the school self-evaluation (SSE) system provides "some degree of constraint on otherwise autonomous school leaders".[124]
- Comparative research on SSEs across four jurisdictions (Ireland, New Zealand, Pakistan and Dubai) suggests that self-evaluation plays out differently depending on context. The degree of centralisation, hierarchy and control may be particularly significant. The study concludes that to be effective, SSE needs to be obligatory; there needs to be a focus on teaching and learning processes – not just outputs; and training needs to build capacity among professionals at all levels (including teachers, school leaders and those in school support roles).[125]
- In the Netherlands, research on a 'principle-based approach' to inspection based on general, "broadly stated rules or 'principles'" with school boards that are assumed to be intrinsically motivated suggests that this approach can help build professional pride and self-confidence, feeding a culture of continuous improvement.[126]

[120] Dandala 2019 p8
[121] Jisu 2020, Hwa 2021
[122] Jisu 2020 p.v
[123] McNamara et al, 2020; McNamara et al. 2021; Gardezi 2024;
[124] Gardezi 2024 p11
[125] Gardezi 2024 p12
[126] Honingh et al, 2022 p238

However, this may depend on an absence of societal pressure and an existing sense that education is of an acceptable quality.[127]

Evidence on the relationship between accountability and teacher turnover is mixed. 'Turnover effects' encompass all three forms of turnover: "'attrition' (where teachers depart the profession), 'between-school churn' (where teachers move from one school to another) and 'within-school churn' (where teachers move from one group of students to another within the same school)".[128] Several of the reviewed US NCLB studies reported an impact on teacher turnover, and one found no effect.[129] Patterns differed, with two studies finding that exit rates increased in lower-rated schools[130]. In contrast, one study found that teacher exit rates were lower in schools in the lower rating band.[131] A fourth focused on tested year groups in New York, on the grounds that teachers working with these pupils would experience the most intense accountability pressure. Contrary to expectations, turnover in these year groups was in fact reduced, partly due to a decrease in 'within school churn'. The authors conclude that "schools are keeping their teachers in the testing grades at a higher rate" and note that this apparent prioritisation of tested year groups may have a negative impact elsewhere.[132] These findings therefore represent another form of 'strategic behaviour' in response to test-based accountability. Another study took a district-level perspective, showing that local governments responded to accountability pressures by increasing teacher pay, changing policies related to teacher promotions and transfers, and by replacing teachers in economically disadvantaged students' schools at higher rates.[133]

Turnover has also been linked to inspection pressure, though the evidence is not unequivocal. Several studies examined the impact of Ofsted inspections in England. One found no increase in turnover following a negative inspection,[134] but a second study found that 'stuck schools' – those that experienced two negative inspection gradings in a row – subsequently faced greater turnover.[135] One theoretical, rather than empirical paper noted that schools that do well in inspections often find it easier to attract better teachers – raising questions with regard to equity.[136]

Where turnover does occur, its effects on pupils will depend on who replaces the departing teachers. Three of the above US studies explore this question, finding that teacher departures do not result in replacement by less experienced or less qualified

[127] Ehren 2025, p53
[128] Menzies (2023) p3
[129] Eren et al (2023)
[130] Clotfelter et al. (2004), Feng et al. (2018)
[131] Dizon-Ross (2020)
[132] Boyd et al. (2008) p108
[133] Gregg and Lavertu (2023) p3
[134] Hussain, 2015
[135] Ehren et al. (2023); Muñoz-Chereau et al. 2022
[136] Tian 2025

teachers. Indeed two studies found that schools replaced departing teachers with more qualified, more experienced, or more able teachers.[137] Whereas these US studies tend to focus on workforce patterns within a specific year group, or in schools receiving a low rating, a Norwegian study looked at the introduction of a wholly new accountability system. The authors find that the new system resulted in a considerable increase in teacher attrition, and that this effect was concentrated among 'high ability' teachers. However, they find that these teachers were replaced by high ability teachers and that the 'quality' of teachers (as measured by university GPA/grades) increased overall.[138] It is worth noting that, in this case, the accountability system in question was said to be a 'low stakes' one.

Turnover therefore represents both an effect in itself, and potential explanatory mechanism for improvements in pupil outcomes. In other words, if accountability leads to increases in turnover and these drive improvements in 'teacher quality', then this might explain some of test-based accountability's positive effects on pupil attainment. However, this mechanism would only be viable in circumstances where there is a willing supply of well-qualified teachers, so it is not necessarily a finding that travels well across contexts.

Pupil distribution

One of accountability's main documented effects relates to the distribution of pupils. Competition in market-based accountability systems tends to drive pupils towards higher performing schools or settings. These sorting effects are most marked when there is more choice available and competition is greater (whether due to school proximity, or within the private sector).[139] A 'sorting to better schools' effect might be considered desirable by policy makers seeking to use 'laissez faire' mechanisms to drive up standards, since this may ensure more pupils are taught in high-performing schools.

Sorting may have negative effects on communities and on pupils who remain in lower-performing schools – and this may be harmful to equity. Pupil sorting has been noted in studies of England and Wales (though it is worth noting that the Welsh study was based on interviews with school leaders in just four schools).[140] Two studies in England demonstrate that choice and sorting drives socio-economic segregation – with socially disadvantaged students becoming more concentrated in low-rated schools. One of these studies used census data, from 1981 to 2011 to identify effects[141] and the other was based on the aforementioned analysis of 'stuck schools.[142] These sorting effects may drive inequitable outcomes.

[137] Clotfelter et al. (2004); Boyd et al. (2008) Dizon-Ross (2020)
[138] Gjefson and Gunnes (2020)
[139] Bassok et al. (2019) Nunes (2015) Morozumi and Tanaka, (2023)
[140] Evans (2023).
[141] McArthur and Reeves (2022)
[142] Munoz Chereau et al., 2022

Conclusion

Rather than accountability being linked to a specific effect, school districts, leaders and teachers respond to the incentives that are built into the system and this response is conditioned by context. Where accountability systems create pressure, there is a response. Where pressure is greater, the response tends to be greater, and when the incentives change – for example when a test metric shifts – responses shift too.

It is therefore unsurprising that on balance, performance-based accountability systems often have a positive effect on measured pupil attainment. These effects appear to be particularly concentrated in lower-performing schools and systems, and potentially among lower attaining pupils. This may be beneficial for educational equity. There is also some evidence that these effects are more marked when measures are higher profile, more tightly focused, and when they are based on standardised measures. However these improvements may be limited to what is measured and there may be wider – potentially negative repercussions elsewhere.

A much smaller body of evidence from the US points to positive long run effects on a range of life outcomes. These include employment, welfare dependency, and criminality. These follow plausibly from educational achievement, or 'human capital accumulation'.

Performance-based accountability's effects on attainment may be explained by:

1. **Heightened pressure that drives schools to focus more sharply on certain subjects, tasks or content.** Pressure might come either from hierarchical authorities or parental choice and community pressure.
2. **Pressure leading to changes in staffing and replacement with more effective colleagues.** However, this is only viable where teachers can be replaced with other high-performing professionals.
3. **Pupils may move to higher performing schools in market-based systems**. However, this may drive segregation and could have inequitable impacts.

The desirability of some of these effects is contestable. Teacher and pupil mobility, and the degree of emphasis on test-related content can be desirable or undesirable depending on one's perspective. Policy makers therefore have to be careful to align their chosen approach to accountability with their policy intent.

Additionally, there is extensive evidence of more unequivocally undesirable effects, such as exclusion of lower-performing pupils and cheating. These effects may be driving some attainment gains and it is at least possible that in their absence, positive attainment effects would be reduced.

Research on other, more horizontal forms of accountability often presents more appealing conclusions but is often (though not always) smaller scale in nature. There is some evidence that different inspection arrangements, including those that are less hierarchical or performance-focused may have a positive effect on collaboration and improvement, while mitigating some of the negative effects associated with other approaches, as well as being less anxiety and stress-inducing.

Whereas hierarchical approaches to accountability can be helpful to policy-makers as a policy lever – by driving alignment across the sector – other, more horizontal forms of accountability may promote more innovation and collaboration. Once again, chosen approaches need to align with intent: policy makers have relatively few levers with which to shape the education system, and one of accountability's functions is 'evaluation and control'. Thus, alignment might be an intended consequence. On the other hand, policy-makers may want to empower frontline staff so that they can innovate and solve problems – in which case overly hierarchical forms of accountability might be maladaptive. Policy-makers may therefore want to opt for different approaches depending on their goals and priorities.

Most accountability systems do not just choose one approach to accountability, instead, they combine hierarchical, market, horizontal, network-based and participative approaches. The popularity of risk-based or differentiated inspections suggests that many systems have either explicitly or implicitly chosen to deploy more hierarchical methods as a 'safety net' to bring up the bottom[143], in combination with other mechanisms to drive continuous improvement beyond this.

We return to the question of how systems combine different accountability elements – and the effects this might have – in the next sections of this report.

[143] As noted in 'Purpose and Structure" efficiency is also part of the rationale

Part 4 References

NB: References for policy documents and additional sources can be found in the full and underlying reports and tables in the CES library on the CES website.

Altrichter, H., & Kemethofer, D. (2015). Does accountability pressure through school inspections promote school improvement? *School Effectiveness and School Improvement*, *26*(1), 32–56. https://doi.org/10.1080/09243453.2014.927369

Anderson, P. M., Butcher, K. F., & Schanzenbach, D. W. (2017). Adequate (or adipose?) yearly progress: assessing the effect of "No Child Left Behind" on children's obesity. *Education Finance and Policy*, *12*(1), 54–76.

Andrabi, T., Das, J., & Khwaja, A. I. (2017). Report cards: The impact of providing school and child test scores on educational markets. *American Economic Review*, *107*(6), 1535–1563.

Atchison, Drew, Le Floch, Kerstin Carlson, Özek, Umut, Luu, Justin, Hurlburt, Steven, & Boyle, Andrea. (2025). The Effects of School Accountability Ratings in California Under ESSA on Student Graduation and College and Career Readiness. *AERA Open*, *11*, 23328584251349184. https://doi.org/10.1177/23328584251349186

Bassok, D., Markowitz, A. J., Bellows, L., & Sadowski, K. (2021). New evidence on teacher turnover in early childhood. *Educational Evaluation and Policy Analysis*, *43*(1), 172–180.

Bergbauer, A. B., Hanushek, E. A., & Woessmann, L. (2024). Testing. *Journal of Human Resources*, *59*(2), 349–388.

Bertoni, M., Brunello, G., & Rocco, L. (2013). When the cat is near, the mice won't play: The effect of external examiners in Italian schools. *Journal of Public Economics*, *104*, 65–77.

Bokhari, F. A. S., & Schneider, H. (2011). School accountability laws and the consumption of psychostimulants. *Journal of Health Economics*, *30*(2), 355–372.

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2008). The impact of assessment and accountability on teacher recruitment and retention: Are there unintended consequences? *Public Finance Review*, *36*(1), 88–111.

Brown, M., McNamara, G., O'Hara, J., & Shevlin, P. (2020). Polycentric inspection: A catalyst for sparking positive interactions in educational networks. *Evaluation*, *26*(1), 76–97.

Burgess, S. (2025). *A review of quantitative causal research evidence on school accountability: updating the evidence*.

Burgess, S., Wilson, D., & Worth, J. (2013). A natural experiment in school accountability: The impact of school performance information on pupil progress. *Journal of Public Economics*, *106*, 57–67.

Burns, T., Köster, F., & Fuster, M. (2016). *Educational research and innovation education governance in action lessons from case studies: Lessons from case studies*. OECD Publishing.

Camargo, B., Camelo, R., Firpo, S., & Ponczek, V. (2018). Information, market incentives, and student performance: Evidence from a regression discontinuity design in Brazil. *Journal of Human Resources*, *53*(2), 414–444.

Chiang, H. (2009). How accountability pressure on failing schools affects student achievement. *Journal of Public Economics*, *93*(9–10), 1045–1057.

Cilliers, J., Mbiti, I. M., & Zeitlin, A. (2021). Can Public Rankings Improve School Performance?: Evidence from a Nationwide Reform in Tanzania. *Journal of Human Resources*, *56*(3), 655–685.

Clapham, A. (2015). Post-fabrication and putting on a show: examining the impact of short notice inspections. *British Educational Research Journal*, *41*(4), 613–628.

Clotfelter, C. T., Ladd, H. F., Vigdor, J. L., & Diaz, R. A. (2004). Do school accountability systems make it more difficult for low-performing schools to attract and retain high-quality teachers? *Journal of Policy Analysis and Management*, *23*(2), 251–271.

Coelli, M., & Foster, G. (2024). Unintended consequences of school accountability reforms: Public versus private schools. *Economics of Education Review*, *99*, 102523.

Dandala, S. (2019). Human resource policy and teacher appraisal in Ontario in the era of professional accountability. *Management in Education*, *33*(1), 5–10.

Dee, T. S. (2020). Learning from the Past: School Accountability before ESSA. *Stanford, CA: The Hoover Education Success Initiative. Retrieved January*, *7*, 2021.

Deming, D. J., Cohodes, S., Jennings, J., & Jencks, C. (2016). School accountability, postsecondary attainment, and earnings. *Review of Economics and Statistics*, *98*(5), 848–862.

Dizon-Ross, R. (2020). How does school accountability affect teachers?: Evidence from New York City. *Journal of Human Resources*, *55*(1), 76–118.

Dyson, L. (2020). Walking on a tightrope: Agency and accountability in practitioner inquiry in New Zealand secondary schools. *Teaching and Teacher Education*, *93*. https://doi.org/10.1016/j.tate.2020.103075

Ehren, M. (2025). *Context, mechanisms and outcomes of accountability systems in fourteen education systems. Working paper for the Centre for Education Systems*.

Ehren, M. C. M. (2025). Context, mechanisms and outcomes of accountability systems in fourteen education systems. In *Working paper for the Centre for Education Systems*.

Ehren, M. C. M., Gustafsson, J.-E., Altrichter, H., Skedsmo, G., Kemethofer, D., & Huber, S. G. (2015). Comparing effects and side effects of different school inspection systems across Europe. *Comparative Education*, *51*(3), 375–400.

Ehren, M., Hutchinson, J., & Muñoz-Chereau, B. (2023). Place-based disparities faced by stuck schools in England: a contextual understanding of low performance and the role of inspection outcomes. *School Effectiveness and School Improvement*, *34*(4), 401–418.

Ehren, M., Perryman, J., & Shackleton, N. (2015). Setting expectations for good education: How Dutch school inspections drive improvement. *School Effectiveness and School Improvement*, *26*(2), 296–327.

Eizadirad, A., Martinez, X., & Ruminot, C. (2016). Comparative analysis of educational systems of accountability and quality of education in Ontario, Canada and Chile: Standardized testing and its role in perpetuation of educational inequity. *Interfaces Brasil/Canadá*, *16*(2), 54–88.

Eren, O., Figlio, D., Mocan, N., & Ozturk, O. (2022). *School Accountability, Long-Run Criminal Activity, and Self-Su¢ ciency*.

Evans, G. (2023a). Scratching beneath the surface: using policy archaeology to better understand education reform in Wales. *Management in Education*, 08920206231160646.

Evans, G. (2023b). *Understanding the impact of school inspection: An analysis of how school leaders in Wales interpret and respond to inspection outcomes*. University of Wales Trinity Saint David (United Kingdom).

Feng, L., Figlio, D., & Sass, T. (2018). School accountability and teacher mobility. *Journal of Urban Economics*, *103*, 1–17.

Figlio, D., & Loeb, S. (2011). School accountability. *Handbook of the Economics of Education*, *3*, 383–421.

Figlio, D. N., & Getzler, L. S. (2006). Accountability, ability and disability: Gaming the system? In *Improving school accountability* (Vol. 14, pp. 35–49). Emerald Group Publishing Limited.

Figlio, D. N., & Rouse, C. E. (2006). Do accountability and voucher threats improve low-performing schools? *Journal of Public Economics*, *90*(1–2), 239–255.

Figlio, D. N., & Winicki, J. (2005). Food for thought: the effects of school accountability plans on school nutrition. *Journal of Public Economics*, *89*(2–3), 381–394.

Gardezi, S. (2024). The development and status of self-evaluation as a component of school inspection and quality assurance: A comparison across four countries. *Heliyon*, *10*(20). https://doi.org/10.1016/j.heliyon.2024.e39321

Gjefsen, H. M., & Gunnes, T. (2020). How school accountability affects teacher mobility and teacher sorting. *Education Economics*, *28*(5), 455–473.

Gregg, J. J., & Lavertu, S. (2023). Test-based accountability and educational equity: Breaking through local district politics? *Economics of Education Review*, *97*, 102485. https://doi.org/https://doi.org/10.1016/j.econedurev.2023.102485

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management*, *24*(2), 297–327.

Hanushek, E. A., Saenz-Armstrong, P., & Salazar, A. (2024). *Balancing Federalism: The Impact of Decentralizing School Accountability*. http://www.nber.org/papers/w32351

Hargreaves, A. (2020). Large-scale assessments and their effects: The case of mid-stakes tests in Ontario. *Journal of Educational Change*, *21*(3), 393–420.

Holbein, J. B., & Ladd, H. F. (2017). Accountability pressure: Regression discontinuity estimates of how No Child Left Behind influenced student behavior. *Economics of Education Review*, *58*, 55–67.

Honingh, M., van Genugten, M., de Gooyert, V., & Blom, R. (2022). What is the impact of a principle-based inspection approach? *School Effectiveness and School Improvement*, *33*(2), 236–259.

Hussain, I. (2015). Subjective performance evaluation in the public sector: Evidence from school inspections. *Journal of Human Resources*, *50*(1), 189–221.

Hwa, Y.-Y. (2021). Contrasting approaches, comparable efficacy?: How macro-level trust influences teacher accountability in Finland and Singapore. In *Trust, Accountability and Capacity in Education System Reform*. Taylor & Francis.

Jacob, B. A., & Levitt, S. D. (2003). *Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating*.

Jang, E. E., & Sinclair, J. (2018). Ontario's educational assessment policy and practice: a double-edged sword? *Assessment in Education: Principles, Policy & Practice*, *25*(6), 655–677. https://doi.org/10.1080/0969594X.2017.1329705

Jisu, R. (2020). *Supporting Professional Capacity of Teachers Through Teacher Evaluation (Doctoral dissertation)*.

Jones, K. L., Tymms, P., Kemethofer, D., O'Hara, J., McNamara, G., Huber, S., Myrberg, E., Skedsmo, G., & Greger, D. (2017). 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*, *43*(6), 805–822.

Kay, A. (2012). Policy trajectories and legacies: path dependency revisited. In *Routledge handbook of public policy* (pp. 480–490). Routledge.

Keane, A. (2023). Changing features of school inspection: a perspective from Wales. *Forum*, *65*(3), 81–90.

Kim, K. (2024). *Teacher Decision-Making and Classroom Autonomy Under Test-Based Accountability: A Comparative Analysis*. The Florida State University.

Klimecki, R., & Wilson, D. (2025). Decentering governance models: education policy and school accountability in Post-Devolution Wales. *Critical Policy Studies*, *19*(3), 456–476.

Lee, Y.-J. (2024). The discursive construction of responsibility among teachers in Singapore: "Everything is good, so we do everything." *Asia Pacific Education Review*, *25*(1), 187–200.

Mansfield, J., & Slichter, D. (2026). The long-run effects of consequential school accountability. *Journal of Labor Economics*, *44*(1), 0.

Maroy, C., & Pons, X. (2021). Re-conceptualising education policy trajectories in a globalised world: lessons from a multi-level comparison of accountability in France and Quebec. *Comparative Education*, *57*(4), 560–578.

McArthur, D., & Reeves, A. (2022). The unintended consequences of quantifying quality: Does ranking school performance shape the geographical concentration of advantage? *American Journal of Sociology*, *128*(2), 515–551.

McClurg, K., Cantley, I., & Donnelly, C. (2025). The possibilities and problematics of school evaluation: post-primary teachers' perspectives on the impact of school inspections in a high stakes inspection system. *Irish Educational Studies*, *44*(3), 421–443.

McElroy, K. (2023). Does test-based accountability improve more than just test scores? *Economics of Education Review*, *94*, 102381.

McNamara, G., O'Hara, J., Brown, M., & Quinn, I. (2020). Quality assurance in Irish schools: Inspection and school self-evaluation. *Administration*, *68*(4), 161–180.

McNamara, Gerry, Brown, Martin, Gardezi, Sarah, O'Hara, Joe, O'Brien, Shivaun, & Skerritt, Craig. (2021). Embedding Self-Evaluation in School Routines. *Sage Open*, *11*(4), 21582440211052550. https://doi.org/10.1177/21582440211052552

Menzies, L. (2023). Continuity and Churn: Understanding and responding to the impact of teacher turnover. *London Review of Education*. https://doi.org/10.14324/LRE.21.1.20

Menzies, L., Yates, W., & Huband-Thompson, B. (2023). *Balancing Act: Navigating the Tensions in our School System*. https://ippr-org.files.svdcdn.com/production/Downloads/balancing-act-school-system-august-23.pdf

Millard, W., Small, I., & Menzies, L. (2017). *Testing the Water: How assessment can underpin, not undermine, great teaching*. The Centre for Education and Youth / Pearson. https://www.researchgate.net/publication/330812163_Testing_the_Water_How_assessment_can_underpin_not_undermine_great_teaching#fullTextFileContent

Møller, J., Hall, D., & Normand, R. (2025). Public education and teacher professionalism in an age of accountability. *Critical Studies in Education*, *66*(5), 664–682.

Moreton, H. J., Boylan, M., & Simkins, T. (2017). Headteachers who also inspect: Practitioner inspectors in England. In *School inspectors: Policy implementers, policy shapers in national policy contexts* (pp. 137–158). Springer.

Morozumi, A., & Tanaka, R. (2023). *School Accountability and Student Achievement: Neighboring Schools Matter*. Research Institute of Economy, Trade and Industry (RIETI).

Munoz Chereau, B., Hutchinson, J., & Ehren, M. (2022). *"Stuck" schools: Can below good Ofsted inspections prevent sustainable improvement?*

Nunes, L. C., Reis, A. B., & Seabra, C. (2015). The publication of school rankings: A step toward increased accountability? *Economics of Education Review*, *49*, 15–23.

O'Mara-Eves, A., Vigurs, C., Ko, S. Y. (Jessica), & Simon, A. (2025). *Rapid systematic evidence review of school accountability policies in secondary education across fourteen jurisdictions*. EPPI Centre.

Pinto, L. E. (2016). Tensions and fissures: The politics of standardised testing and accountability in Ontario, 1995–2015. *The Curriculum Journal*, *27*(1), 95–112.

Reback, R., Rockoff, J., & Schwartz, H. L. (2014). Under pressure: Job security, resource allocation, and productivity in schools under No Child Left Behind. *American Economic Journal: Economic Policy*, *6*(3), 207–241.

Rezai-Rashti, G. M., & Segeren, A. (2023). The game of accountability: perspectives of urban school leaders on standardized testing in Ontario and British Columbia, Canada. *International Journal of Leadership in Education*, *26*(2), 260–277.

Rockoff, J., & Turner, L. J. (2010). Short-run impacts of accountability on school quality. *American Economic Journal: Economic Policy*, *2*(4), 119–147.

Rouse, C. E., Hannaway, J., Goldhaber, D., & Figlio, D. (2013). Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure. *American Economic Journal: Economic Policy*, *5*(2), 251–281.

Şen, A. Y., & Maya, İ. (2022). School quality assurance applications for basic education schools. *Education Quarterly Reviews*, *5*.

Tian, M. (2025). Embracing complexity: rethinking education inspection in England. *Educational Review*, *77*(5), 1500–1516. https://doi.org/10.1080/00131911.2024.2325660

Torres, R. (2021). *Does test-based school accountability have an impact on student achievement and equity in education?: A panel approach using PISA*.

Wang, F. (2017). The Ontario quandary: Principals' perceptions of accountability policy and social justice. *Journal of School Leadership*, *27*(3), 392–420.

Part 5: Overarching Findings

Typical approaches

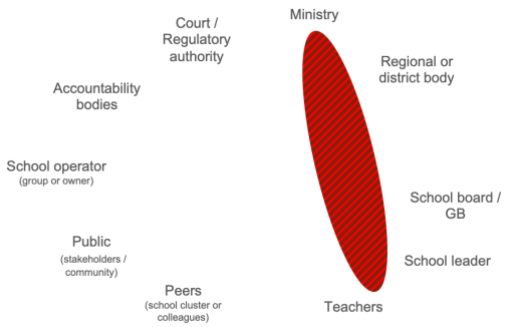
In 'Purposes and Structures' we saw that accountability systems seek to fulfil a number of different functions by bringing together a web of different accountability elements. These involve a range of actors, relations, measures and consequences, which can be characterised with reference to various 'types' and configurations.

The research reviewed in relation to 'Contexts and Effects' suggests that different approaches take their roots in different contexts and are associated with various interconnected effects. These relate to pupil outcomes, school and teacher practices, workforce dynamics, and pupil distribution – and these effects may not always align with their intent.

In this section we look across the characteristics of different accountability systems and link the main approaches to the contexts that may give rise to them, and the circumstances and mechanisms that shape their effects. In order to do so we return to the seven main accountability configurations set out in 'Purposes and Structures' reviewing their component characteristics, how these approaches relate to the contexts in which they are found, and what effects we might expect to see associated with them. We also draw on international benchmarking data – not in order to imply any causality, but to locate our discussion in context.

1. Chain of command

This approach involves actors being accountable through a step-by-step chain – for example from Head, to school board, to regional authority, to national government. This normally involves managerial relations, but contractual and legal- or compliance-based accountability can also play a role.



Chains of command are the most common form of accountability among CES jurisdictions and represent a somewhat traditional, bureaucratic model. Singapore, France, Estonia, Poland and Japan exemplify this approach, but almost all jurisdictions use chains of command to some extent. The jurisdictions that most deviate from this approach are:

- England's multi-academy trust system, which uses contractual relations within its 'Polycentric' approach (although the new 'RISE' programme' may represent a return

to chain-of-command approaches, since regional ministry teams will become more directly involved with schools).
- Finland, which prioritises the 'Horizontal Professional' approach combined with 'Collective system accountability'.

Systems in which accountability is dominated by chains of command tend to see schools as state institutions to be managed, rather than regulated.[144] Systems that lean heavily towards this approach may place less emphasis on autonomy, be less influenced by new public management, and/or be more hierarchical in character. This is somewhat born out in international comparisons based on 'Power Distance Index' – a measure of the acceptance of hierarchy. Countries in bold below represent 'typical cases' of the chain of command, but as noted above, almost all countries deploy this approach to some extent.

Power Distance Index 2

Eng	Wal	NI	Scot	ROI	Ont	NZ	**Sing**	**Jap**	Fin	**Est**	**Pol**	Neth	**Fra**
35	35	35	35	28	39	23	**74**	**54**	33	**40**	**68**	38	**68**

Sources: Pisa Table II.B1.6.1 (Devolved jurisdictions Table II.B2.34 2024); Home Nations based on UK

Countries with a more structured chain of command often align with countries in which policy making is more centralised or bureaucratic. However, new approaches to governance have sometimes been 'layered' on top – as is the case in post-Soviet/communist systems (Estonia and Poland).

Components

The operation of chains of command varies considerably, and can involve a range of measures, tools, and consequences.

Commonly used tools include:
- **Inspection**
- **Reporting –** for example through management information systems
- **Local oversight –** for example through 'cluster superintendents'

Measures can relate to:
- **Inputs** – for example, by auditing financial reports or teacher HR and safeguarding records

[144] Majone (1997)

- **Processes** – for example, through lesson observations that try to gauge the quality of teaching, or by requiring leaders to explain their approach to curriculum design.
- **Outcomes** – such as test results – which may either be public, or used to inform decisions within the chain of command.

Consequences usually involve a mixture of support and capacity building, and sanctions.

- **Additional support or heightened oversight is often the preferred consequence.** Chains of command lend themselves towards this approach because managers the next level up can take a hands-on role, and steps can be taken to escalate. Moreover, chains of command are often underpinned by somewhat state-centric logics – in contrast to polycentric systems (which we return to below).
- **Despite this, almost all systems have a system of sanctions too.** As shown in 'Purpose and Structure', almost all reviewed jurisdictions include rapid reinspection, warning letters, improvement plans, and category designations within their suite of accountability consequences.

Effects

Chains of command operate in various ways and the effects of 'chains of command' will likely depend on the precise mix of tools, measures and consequences, how leaders throughout the chain of command go about their work, and the jurisdiction's particular context.

Where there is more of a focus on sanctions:

- This may set a clear floor on standards and drive up attainment, particularly in low-performing settings, or for low-attaining pupils.[145]
- This may be driven by a sharper focus on specific outcomes or priorities – particularly where measures are performance based[146] – as well as heightened pressure, which can drive turnover and replacement of teachers (potentially by more effective colleagues).[147]
- This pressure may also create an incentive to exclude pupils or cheat in exams,[148] alongside stress and anxiety for the workforce.[149]

[145] Cilliers et al (2021); Deming et al., 2016 ; Feng et al. (2018); Burgess et al. (2013); Bergbauer et al (2024) ; Torres (2021)
[146] Bergbauer et al (2024); Hanushek et al (2024)
[147] Clotfelter et al. (2004); Boyd et al. (2008) Dizon-Ross (2020); Gjefson and Gunnes (2020)
[148] Deming et al (2016); Eren et al (2023), Figlio and Getzler (2006) Gregg and Lavertu (2023) Rezai-Rashti and Segeren (2023); Coelli and Foster (2024); Cilliers et al (2021); Jacob and Levitt (2003); Bertoni et al. (2013); Pinto (2016
[149] McClurg et al, 2025; Evans 2023

- The extent to which sanctions drive pressure might depend on how likely consequences are, and the robustness with which they are implemented. However, these are questions of enactment that go beyond this study's remit.

Where chains of command are less sanction-focused they may prompt 'organisational learning'. This may be more likely when there is an absence of pressure for immediate improvement, support and resources are made available, and there is capacity to engage and learn. [150]

2. Vertical teacher accountability

This approach involves teachers being directly accountable to bodies outside of the school.

Vertical teacher accountability skips the 'chain of command'. The focus can be either on compliance with teaching standards, or on development/professional practice.

<u>Components</u>

Most jurisdictions have arrangements in place for compliance-based professional accountability. This either happens through an independent/partially-independent regulatory body (see section 'Purpose and Structure'), or through the ministry itself.

- Judgements and consequences tend to be responsive to claims of malpractice that are linked to a code of practice.
- Judgements will be made in a quasi-judicial way, and being barred from the profession is the typical consequence.
- Although cases of malpractice tend to be fairly rare, the *possibility* of being held to account can have wider significance.[151]

Development, or quality-focused approaches are more unusual. France, Singapore, and – to some extent – Japan and New Zealand exemplify this approach.

- Measures may be:
    - Input-focused, where teachers have to demonstrate that they have participated in a certain amount of professional development (or reflect on their practice);
    - Process-focused, where they must account for how they are going about their job, and their practice may be observed;
    - Outcome-focused, where their students' performance is tracked.

[150] Ehren (2025)
[151] Mulgan (2000) p567

- Although forums in this approach are by definition external, school leaders may contribute too – as is the case in Singapore and Japan.
- Consequences tend to be employment related (for example linked to salary or promotion).
- Jurisdictions in which teachers are externally evaluated tend to have teachers as state employees. This form of accountability is therefore associated with contexts in which head teachers have less autonomy over teacher appointments – as can be seen in the table below.

Principals Had Main Responsibility for Hiring Teachers (%)

Eng	Wal	NI	Scot	ROI	Ont	NZ	**Sing**	**Jap**	Fin	Est	Pol	Neth	**Fra**
84	48	30	79	N/A	73	74	**20**	**24**	57	94	97	64	**10**

Source: Pisa Table II.B1.6.1 (Devolved jurisdictions Table II.B2.34 2024)

- Like chains of command, jurisdictions that deploy this approach tend to have a more hierarchical culture (see previous table).

Effects

Evidence on the effects of 'compliance-focused' vertical teacher accountability was not found as part of this study. It is perhaps best understood as guarding against egregious malpractice, and as a 'silent sentinel', maintaining a floor on standards.

It is notable that three of the jurisdictions that had previously made most use of quality-focused vertical teacher accountability have stepped back from, or reformed their systems.

- Japan and New Zealand have both softened or abolished their recertification requirements, while Singapore has reformed its performance management system (EPMS) to take a more 'holistic' and developmental approach.
- Reforms have been prompted by concerns about workload or burden for teachers; a desire to ease teacher shortages; and desire to emphasise professional growth.

The effect of quality or development focused external accountability is mixed. There is some positive evidence for impact on teaching and attainment but evidence of stressful and unproductive consequences too.

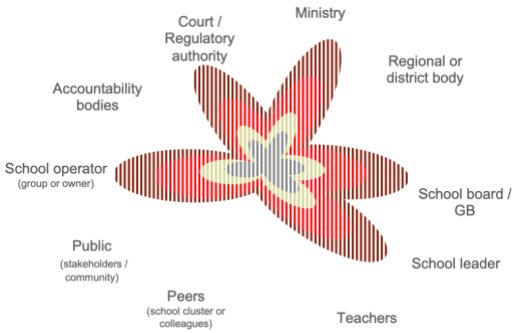
- Singapore's previous EMPS system may have led to improved teaching practices and high student performance, but it is also said to have generated high workload and gaming (according to one small-scale study).[152]

[152] Hwa (2021) p230

- There is some evidence from the US (where most of the large-scale quantitative studies of performance-based accountability have been conducted) that – where teachers are held accountable for student test scores, this can lead to increases in pupil achievement.[153]
- Japan's – now abandoned – recertification scheme was seen to be burdensome and to drive compliance.[154]
- One small-scale comparative study suggests that although teachers in Finland have to report up to their municipality, this has limited day-to-day impact on them compared to in Singapore.[155]
- Another comparative study – this time based on analysis of TALIS data pointed to differences between the effects of teacher evaluation systems in Finland and Japan, once again stressing compliance within Japan's system and suggesting that Finland's system had little impact on practice. The author concluded that teacher evaluation systems were more likely to build professional capacity when "coupled with school-level actions to support their professional growth and a school-level climate of shared responsibility"[156]
- Even where accountability focuses on the institution (governing body or school leader), pressure can still be passed on to teachers.

3. Polycentric accountability

This approach involves multiple intermediary institutions, with legal and contractual forms of accountability often playing an important role, alongside managerial relations.[157]



Polycentric approaches are associated with governance and regulation, rather than state-centric management.[158] Such approaches allow the state to step back from directly providing or managing public services like education.

Actors can include school proprietors or operators. This can be seen in England, Northern Ireland, the Republic of Ireland, the Netherlands and Ontario.

[153] Hanushek et al (2024)
[154] Huang et al 2024
[155] Hwa (2022) p227
[156] Jisu (2020) p.v
[157] The term "polycentric accountability" has also been used to describe certain approaches to inspection and horizontal accountability, however we use the term draws on the fields of governance and policy.
[158] Ball, S. J., & Junemann, C. (2012). *Networks, new governance and education*. Policy Press Bristol.; Hughson and Menzies (2026);

Forums can include independent or quasi-independent bodies. These include independent inspectorates (as is the case in England, Wales, New Zealand, Netherlands) and independent, or quasi-independent teacher regulation bodies (as is the case in Wales, Northern Ireland, Scotland, Ireland, Ontario and New Zealand – but not England).

Polycentric approaches are primarily concentrated in CES' Anglophone jurisdictions.

Polycentricity is commonly related to New Public Management (NPM). The intention behind NPM is for the state to incentivise innovation and improvement by stepping back and moving into a 'monitoring' or 'regulating' role, handing autonomy to the front line while introducing market forces into public services.[159] This typically involves the introduction of private-sector management practices such as performance indicators and targets, along with competition for state contracts (as well as 'choice' for 'consumers' – an approach we return to shortly). However, in recent years there has been increasing concern that such approaches may not be effective in driving the intended improvements and that they can crush the pursuit of competence, autonomy and relationships – which might be better drivers of improvement and innovation.[160]

A second explanation for polycentricity can be identified in several CES jurisdictions – namely religious, linguistic or cultural diversity.

- Allowing different churches to run schools with varying degrees of oversight from the state was historically a means of overcoming historic tensions, making space for religious freedom, and partially integrating a church-led system into the state system. This has fundamentally shaped the system in the Republic of Ireland, Northern Ireland and England.
- Similarly in the Netherlands a 'pacification' compromise was reached in 1917 granting state funding to a largely denominational private education sector.[161]
- In Ontario, multiple school boards can operate schools within the same area, with some boards providing French schools, others English, some Catholic, and some secular.
- The case of New Zealand's Charter, or partnership schools is somewhat complicated: the schools were ostensibly created on market-based grounds, but often became a means by which some Māori communities could create schools with an alternative educational approach.[162]

[159] Karsten 1999 p310; Majone 1997
[160] Menzies and Quilter Pinner 2023 p8
[161] Karsten 1999
[162] In 2011 a coalition deal created 'partnership schools', but by 2018 these had been reintegrated into the government sector. A new charter school policy was then introduced in 2024 as part of a new coalition deal.
.

These structural features of education systems can create conditions in which alternative approaches to accountability are developed. This may be one explanation for the differences seen in these jurisdictions, compared to more state-centric 'chain of command' dominated systems

Even in systems that do not lean towards polycentric accountability overall, there are sometimes enclaves of publicly-funded private schools. These sometimes exist for historic or religious reasons and are often held to account through more typically polycentric approaches.

Students educated in government-dependent 'private' schools (%)

Eng	Wal	NI	Scot	ROI	Ont	NZ	Sing	Jap	Fin	Est	Pol	Neth	Fra
59	N/A	N/A	N/A	N/A	N/A	0	0	2	5	5	4	42	7

Source: Table II.B1.6.13 Pisa 2022 – figure for England is 'UK wide' but we include in England as this represents the most pupils – true figure likely to be higher as this largely represents academy schools.

Polycentric approaches are often combined with elements of the 'chain of command'. For example, school operators/proprietors might be inspected by an independent inspectorate, but the ministry might still manage consequences.

Polycentric accountability can often align with less centralised approaches to policy making. This can be seen in Ontario where policy making is often devolved to school boards.

Components

The arrangements for delivering education described above require an adapted set of accountability tools. Whereas in chains of command the state can play a fairly hands-on role, polycentric systems tend to monitor and regulate from a distance.

Polycentric accountability can involve input, process, and/or outcome measurement, but NPM is particularly associated with outcome monitoring.

Accountability tools can include:

- **Collecting compliance information**. This is a widespread approach when managing *inputs* (such as minimum standards) within private-sector 'enclaves'.
- **Sending in inspectors rather than managers**. These inspectors will often evaluate *processes*.

- **Using standardised testing.** This is often the preferred means of monitoring *outcomes* while – at least in theory – giving professionals autonomy over how they deliver their outcomes.

Polycentric systems deploy many of the same consequences as chains of command, but there are some divergences:

- **Consequences can often be relatively 'hands off'.** The state may be unwilling, or unable to involve itself in improvement work. This is accentuated in systems that prioritise market-driven logics (to which we return shortly).
- **Support may come from a school proprietor or operator.** This is the case with school boards in Ontario and multi-academy trusts in England (and Dioceses in religious schools). In the Netherlands, the state provides subsidised access to support packages.
- **Contract management plays a role in consequences and this can include the potential for terminating funding or a licence to operate.** This can be seen in private-sector enclaves, the academy sector in England, and the Charter school system in New Zealand'.[163]

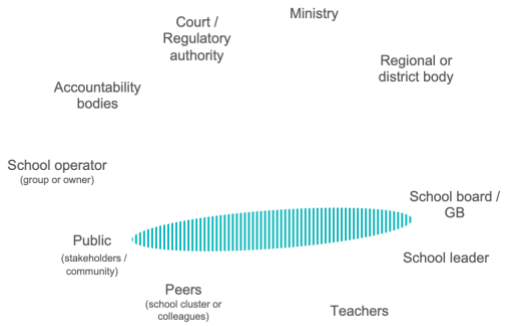
Effects

Part of the rationale for polycentric approaches is that responsibility and autonomy can be shifted to practitioners – or to school leaders and school operators. The World Bank describes this approach as 'School Based Management'.[164] However, in practice, if actors must comply with tightly defined contract requirements and regulations, and their performance is closely tied to consequences, they may not in fact feel particularly autonomous.

Polycentric accountability's effects will vary depending on its components. This includes the type of metrics used (and the incentives this creates); the degree of pressure; and the type of support that is available. These components' effects are discussed above in relation to 'chains of command' and 'vertical teacher accountability'.

4. Market accountability

This approach involves the public holding school leaders or governing bodies to account through market/choice-based approaches



Market accountability, like polycentric accountability, shifts authority away from the

[163] Note that the OECD classifies academies and charter schools as government funded private schools.
[164] World Bank 2008

state. Participatory and horizontal accountability do this too, but in market accountability, judgement rests with parents or pupils and consequences are expected to follow through the 'invisible hand'[165] of aggregated choices.

Market accountability was initially, and to some extent still is, associated with Anglo and Anglo-influenced jurisdictions. During the 1980s these jurisdictions were heavily influenced by 'neoliberal' ideas, linked to thinkers like Milton Friedman and Friederk Hayek[166]. These ideas have since diffused globally.[167] Similarly to the arguments noted above in relation to NPM, the idea is for the state to step back and use market mechanisms that might incentivise efficiency and innovation. However, some countries – including (among CES' jurisdictions), France, Japan and Finland – have been less willing to adopt these approaches.[168]

Some systems have developed policies that intentionally cultivate market forces. These include:

- **Policies that maximise the range of choices available (at least in principle)**. For example by allowing pupils to attend schools other than their nearest school; or, by publicly funding private schools, 'free schools', or 'charter schools'. In its most extreme or purest form (depending on one's point of view), market-based approaches involve voucher systems like those found in Sweden and Chile, as well as in England's (albeit more heavily regulated) early years childcare sector.
- **Policies that maximise the information available to parents.** For example, when the government actively and routinely publishes graded inspection reports and pupil performance data at a school level – as is the case in England, Scotland, New Zealand and the Netherlands.
- **'School Based Management' policies such as linking education funding to pupils.[169]**

A degree of competition is reported in many jurisdictions, however France, Estonia and Finland stand out as being the least competitive. In some systems competition is concentrated at particular levels of the school system – for example, in Japan competition is mainly an important feature of the upper secondary system.

No other schools compete for our students (low scores represent high competition)

[165] Smith [1776] 2005 p364
[166] Hayek (2001); Friedman (2002).
[167] Back in the 1980s, Maurice Kogan noted that arguments about educational accountability in Britain remained "fairly conventional" and that it was American cases that provided examples "which go to the extremes of popular control", showing how one might move "beyond systems of public control altogether" (1986, p51).
[168] Sahlberg (2016), p131
[169] World Bank (2008)

Eng	Wal	NI	Scot	ROI	Ont	NZ	Sing	Jap	Fin	Est	Pol	Neth	Fra
8	N/A	N/A	N/A	13	N/A	11	2	6	62	22	9	7	37

Source: Table II.B1.6.9 Pisa 2022 Table II.B1.6.13 pisa 2022 – figure for England is 'UK wide' but we include in England as this represents the most pupils – true figure likely to be higher.

Components

Measures play a central role in market accountability because they provide the information that is assumed to drive choice. Measures and tools used in market accountability include:

- Outcomes of standardised pupil assessments.
- School ratings (which are themselves derived either from inspections, or pupil assessments – or a combination of the two).

While inspection is a feature of most approaches to accountability, in market-based accountability there is a particular emphasis on informing parents, rather than simply providing formative evaluation. In a chain of command, inspection reports provide information to the school and to superiors in support of improvement work. These reports may not be made public (as is the case in France and Singapore). In contrast, market accountability requires inspection information to be readily digestible by parents so that they can draw on it to make choices. That said, hybridisation means that even in systems that lean towards market accountability, inspection systems typically have to triangulate between multiple purposes.[170]

No individual forum administers consequences in the market-based approach. However, in order for choice-based mechanisms to 'bite', funding tends to follow pupils on a per-capita basis.

Effects

Studies reviewed in 'Contexts and Effects' suggest that market mechanisms may reduce the number of pupils taught in poorly performing settings, and increase pupil attainment.[171] Increases in attainment may be a consequence of pressure and incentives, and/or pupils moving to higher performing schools.

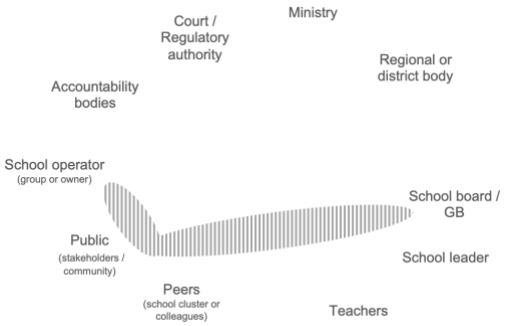
Market accountability is depends on school-to-school competition and this may inhibit collaboration and organisational learning.[172]

[170] Menzies and Quilter-Pinner (2023), p19
[171] Nunes (2015); Bassok et al. (2019); Morozumi and Tanaka, (2023)
[172] Burns et al. (2016)

School choice can drive socio-economic segregation, with socially disadvantaged students becoming more concentrated in low-rated schools.[173] This may follow from more affluent parents being better placed to – and more inclined to – navigate choice-making systems, as well as them having more resources with which to secure access to desirable placements (whether by moving into a different catchment area, or by paying additional fees in partially state-funded private schools). These mechanisms may drive inequitable educational outcomes and entrench social immobility.[174]

5. Participatory accountability to stakeholders

This approach involves stakeholders,(including parents, staff and the local community) participating directly in holding actors to account.



Most systems include bodies through which parents can have a voice in school management. Whereas some of these have a largely advisory or consultative role, others involve parents calling school leaders to account, or reaching a judgement on their performance.

While consultative and advisory staff representation is widespread, it seems rarer for this to constitute an accountability relation, beyond a designated number of places on a governing body/school board. Where staff are directly involved in holding managers to account (through staff councils, or teacher governor roles), this reverses hierarchies. There are established national arrangements for students to play this type of role in France and Estonia (particularly at secondary level).

Participatory accountability is underpinned by differing rationales in different jurisdictions.

- **Participatory accountability can be linked to 'School Based Management'.** This is the case in New Zealand, where parent-dominated school boards were introduced as part of 'Tomorrow's Schools', a policy that sought to reform school management and improve cost effectiveness, as well as in England's school reforms in the late 1980s.
- **Participatory accountability can be linked to ideas of democratic participation and 'civic republicanism'.** For example, in France, accountability is linked to democratic ideals within its republican tradition (which is also linked to ideals around professional accountability).

[173] McArthur and Reeves (2022), Ehren, Hutchinson, & Muñoz-Chereau, (2023)
[174] Ball et al. (1996) p110

Components

Representative bodies are one of the main approaches to participatory accountability. This is particularly marked In New Zealand where parents are often the largest group on school boards. Some jurisdictions, like Estonia, have arrangements whereby individuals from an advisory or consultative parent, teacher, and/or school council are appointed to the board of trustees (or equivalent).

Students, parents and staff surveys often inform self-evaluation. Surveys can also be part of the external inspection process – however, this is perhaps best understood as a way of integrating participatory approaches within other forms of accountability through consultation.

Participatory accountability can extend towards electoral or political accountability when the community is able to vote for actors in school management roles. This approach is widespread in the USA, but among the CES jurisdictions only Ontario takes this approach – through its elected school boards. Ontario also elects students and first nation representatives to its school boards.

There is no typical 'measure', or 'consequence' associated with participatory accountability.

- Where stakeholders are involved in school self-evaluation, the evaluation framework that underpins this may constitute a form of 'measure'.
- Where stakeholders have a say in the appointment of school leaders the consequences can relate to employment .

Effects

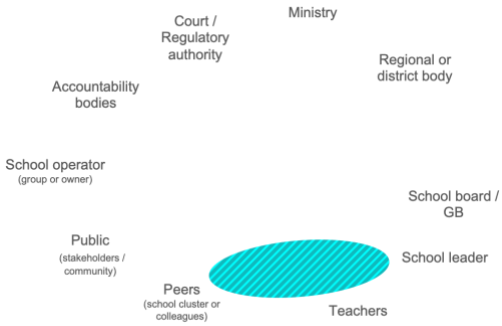
We did not find evidence on the effects of involving stakeholders in governance structures through representative bodies or elections.

Evidence on involving stakeholders in self-evaluation tends to show effects on school and teacher practices, rather than direct impact on pupil performance. These effects include collaboration and teamwork, and use of data for school planning. There is also a potential link between the legitimacy stakeholders lend to the accountability process, and schools' likelihood of taking improvement action in response.[175]

[175] Ehren et al (2013); McNamara et al (2020) p175

6. Horizontal professional accountability

This approach involves teachers being held accountable for their professional practice by their peers, either internally or through networks and clusters.



Like vertical teacher accountability, the *actor* in this approach is the teacher, rather than the institution, however in the horizontal model, the *forum* is a peer, or group of peers, rather than a 'superior' authority. These peers may be internal to the actors' school, or brought together through a network.[176]

Horizontal professional accountability is closely linked to professional development. This can make it hard to discern where structures have an accountability function, and where they are purely professional development-focused. This distinction is also wrapped up in contested definitions of accountability.

Horizontal accountability is sometimes associated with high-trust and egalitarian contexts, but it is not unique to them. Structures linked to horizontal professional accountability can be found across a range of contexts but Finland is often considered an archetype of horizontal professional accountability. It combines this approach with 'Collective system-level accountability' in a non-hierarchical society.,

<u>Components</u>

Horizontal professional accountability typically involves:

- **Teachers and/or leaders coming together regularly to reflect on practice in structured settings.** Wales, Northern Ireland, Scotland, New Zealand, Singapore, Japan, and Finland all have peer-to-peer learning structures and processes that could be argued to have an accountability role.
- **Actors (peers) working with the forum (the teacher, or a network) to reach a shared judgement.**
- **An absence of formal 'measures', combined with judgements that are formative rather than summative.**

However, approaches to horizontal professional accountability are highly heterogeneous. Aspects of horizontal accountability are often embedded within hierarchical accountability structures, for example through:

[176] We combine West et al.'s 'network' accountability type, and a component of their 'professional' accountability types in this approach (2011)

- **Self-evaluation as a component of inspection (usually combined with external validation.)** This can be found to varying degrees within most CES jurisdictions (as described in 'Purpose and Structure') and can involve professional peers reflecting together to reach judgements.
- **Fellow professionals taking part in accountability activities as 'peer inspectors'.** However, it could be argued that when a professional becomes an 'inspector' (even for the day) they take on a hierarchical role and are therefore no longer 'just' a peer.
- **Area or network-level inspection.** Various approaches to this have been described in the Netherlands, Northern Ireland and England[177].
- **Peers can feed into each other's performance appraisals.** This approach is fairly widespread but its prevalence also varies considerably. In some jurisdictions, it seems decisions on promotions and career progression can include peer input. References to this approach were found in documentation from Poland.

Peers involved in teacher appraisals

NZ	Sing	Jap	Fin	Est	Neth	Fra
75	19	43	2	38	10	9

Source: TALIS 2018 Results (Volume II) Annex C Ch3 Table II.3.30 "Percentage of teachers whose school principals report that their teachers are formally appraised at least once a year by the following sources of appraisal (other teachers)". Data unavailable for Poland and ROI (as well as sub-national jurisdictions).

Horizontal professional accountability is therefore perhaps best viewed as a continuum. It involves a collection of components that can serve to make other accountability elements *less* vertical.

[177] Ehren et al (2017) (note the use of 'polycentric' here to denote multiple actors being inspected); Brown et al (2020)

Effects

Horizontal professional accountability may be less stressful and more conducive to collaboration than forms of accountability in which judgements are made by a hierarchical authority which imposes consequences. Although statistics are available from the OECD for levels of collegiality, stress and teacher job satisfaction for half of CES' fourteen jurisdictions, levels are broadly similar – with the exception of Japan and France where collegiality and job satisfaction are somewhat lower.

	NZ	Sing	Jap	Fin	Est	Pol	Neth	Fra
Job satisfaction [1]	85	87	79	86	93	88	94	79
Stress index [2]	10	10	10	9	10	10	9	9
Teacher collegiality [3]	90	94	77	88	90	N/A	92	81

Sources:
1: "All in all, I am satisfied with my job" Annex C Table 2.6 Talis 2024
2: Annex C Table 5 TALIS 2024
3: Agree or strongly agree that teachers can rely on each other. TALIS 2024 Table
Data unavailable for Poland and ROI (as well as sub-national jurisdictions).

Some studies touch on specific variants of horizontal professional accountability:

- Initiatives involving collaborative inquiry and development are said to support situational problem solving and innovation.[178]
- Involvement of peers in inspection can potentially help individuals develop their understanding of the system and to practice skills such as data analysis and lesson observation, whilst also ensuring judgements are explained by an empathetic 'broker'.[179]
- Various findings have been reported in relation to self evaluation:
    - Effects seem to differ depending on the frameworks and guidance used; interactions with other accountability approaches; and cultural context – including trust-levels and teacher expertise.[180]
    - Teachers and leaders often report frustration with approaches that are overly structured – or insufficiently tailored to school goals[181], but comparative research also shows that structure may be beneficial.[182] In England, structured 'self-evaluation forms' were abolished on the grounds that they generated workload and bureaucracy.

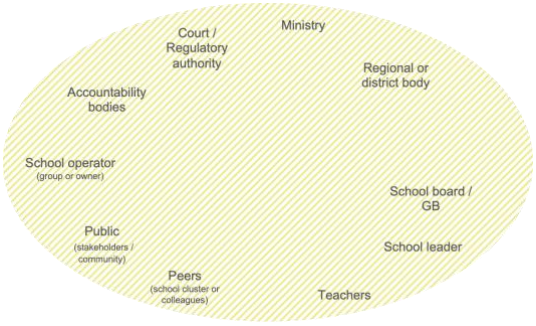
[178] Brown et al (2020); Dyson (2020)
[179] Keane, (2023); Moreton et al (2017)
[180] Hashimoto (2018); Gardezi (2024)
[181] Hashimoto (2018)
[182] Gardezi (2024)

There may be a tension between horizontal mechanisms and more hierarchical approaches.

- Combining self-evaluation with external validation through inspection and linking it to concrete data is believed by some to be beneficial, because it can lend the process legitimacy and equipping leaders to make change.[183]
- External inspection is considered by others to be a counterproductive or a superfluous encroachment on the self evaluation process.[184] In Ireland and Scotland, other forms of accountability have potentially sidelined or undermined self-evaluation.[185]

According to some, performance-based accountability undermined the 'professional inquiry' approach in New Zealand (which may or may not be intended as an accountability tool).186

7. Collective system accountability



This approach involves the education system as a whole being held to account collectively, in order to maximise performance at a jurisdiction level.

This approach does not fall within the scope of this study as set out in Appendix 2 and the Introduction[187]. However, we include it briefly as it plays a role in several systems. **All jurisdictions have systems that monitor system-level performance and feed into policy adjustments.** However, under what we term 'collective system accountability', this information is not just of policy relevance, it is considered a form of shared accountability.

There are similarities between 'collective system accountability' and horizontal professional accountability through clusters or networks (as discussed above). However, by 'collective system accountability' we refer to a jurisdiction-level approach that encompasses all actors within the education system, including teachers, schools, municipalities and government – and potentially even the public.

Finland exemplifies the characteristics of collective system accountability. This is likely linked to the country's context, including social-democratic collectivist norms, low-hierarchy, high trust, a homogenous population, and a highly respected and qualified teaching

[183] McNamara et al. (2020); Gardezi (2024)
[184] Şen & Maya (2022).
[185] McNamara et al. (2022); Proudfoot (2025)
[186] Dyson (2020)
[187] Parameter two states that the actor would be the school organisation or school grouping. However this has already been extended to include teachers.

profession. This may contribute to powerful internalised norms of responsibility towards students, parents, and society among teachers and leaders at every level.[188]

Components

Measures that contribute to collective system accountability include sample-based assessments, thematic inspections, national surveys, international assessments like PISA, longer term destination metrics and research studies. As noted earlier, these exist in most jurisdictions, but in Finland they seem to play a central role in teachers and school's experiences of, and understanding of accountability. In particular, the information from these measures seems to shape judgements that are shared across actors, and changes of practice for everyone. This contrasts with systems in which information from these sources is predominantly used to shape policy or hold governments to account.

Effects

Tools used in collective system level accountability may support data informed planning and targeted intervention.[189]

Relying on collective responsibility might result in poor practice or inconsistent standards going unchallenged, or in improvement recommendations going unactioned.[190]

While a high trust, professionalised environment may be conducive towards collective system accountability, this accountability approach could plausibly strengthen those norms too.

Combining approaches

All jurisdictions combine different accountability approaches. The widespread use of risk-based or differentiated inspection suggests that many jurisdictions target more hierarchical approaches (whether polycentric or chain of command based), at lower performing institutions. Other approaches may then be used to drive improvement above a basic threshold.[191]

Hierarchical or performance-based approaches, and horizontal or process-orientated ones may be complementary, or the combination may lead to incoherence. As will be clear from the earlier discussion of self-evaluation, evidence is unclear in this respect.

[188] Hwa (2022)
[189] Ehren (2025) p20
[190] Hwa (2022); Hwa (2021)
[191] Brown et al. (2020)

- **Performance-based accountability may undermine teacher agency, harming teachers' ability to collaborate and learn.** If horizontal approaches depend on a low threat, high trust environment then the combination of both approaches may be dysfunctional.[192]
- **Different approaches may generate a productive tension, or be mutually reinforcing.** Hierarchical and performance based approaches may provide a safety net; legitimise the case for change; generate a sense of urgency; and highlight inconsistencies.

Layering different approaches to accountability on top of each other, and combining multiple accountability relations may have cumulative effects. For example, one accountability relation might mean a school leader faces consequences from their school proprietor; another might make them subject to consequences from a government body. While each of these relations may be manageable on their own, the combination might lead to more pressure than was intended.

[192] Priestley (2013) p5

Part 5 References

NB: References for policy documents and additional sources can be found in the full and underlying reports and tables in the CES library on the CES website.

Ball, S. J., Bowe, R., & Gewirtz, S. (1996). School choice, social class and distinction: the realization of social advantage in education. *Journal of Education Policy*, *11*(1), 89–112. https://doi.org/10.1080/0268093960110105

Ball, S. J., & Junemann, C. (2012). *Networks, new governance and education*. Policy Press Bristol.

Bassok, D., Markowitz, A. J., Bellows, L., & Sadowski, K. (2021). New evidence on teacher turnover in early childhood. *Educational Evaluation and Policy Analysis*, *43*(1), 172–180.

Bergbauer, A. B., Hanushek, E. A., & Woessmann, L. (2024). Testing. *Journal of Human Resources*, *59*(2), 349–388.

Bertoni, M., Brunello, G., & Rocco, L. (2013). When the cat is near, the mice won't play: The effect of external examiners in Italian schools. *Journal of Public Economics*, *104*, 65–77.

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2008). The impact of assessment and accountability on teacher recruitment and retention: Are there unintended consequences? *Public Finance Review*, *36*(1), 88–111.

Brown, M., McNamara, G., O'Hara, J., & Shevlin, P. (2020). Polycentric inspection: A catalyst for sparking positive interactions in educational networks. *Evaluation*, *26*(1), 76–97.

Burgess, S., Wilson, D., & Worth, J. (2013). A natural experiment in school accountability: The impact of school performance information on pupil progress. *Journal of Public Economics*, *106*, 57–67.

Burgess, S. (2025). *A review of quantitative causal research evidence on school accountability: updating the evidence*.

Burns, T., Köster, F., & Fuster, M. (2016). *Educational research and innovation education governance in action lessons from case studies: Lessons from case studies*. OECD Publishing.

Cilliers, J., Mbiti, I. M., & Zeitlin, A. (2021). Can Public Rankings Improve School Performance?: Evidence from a Nationwide Reform in Tanzania. *Journal of Human Resources*, *56*(3), 655–685.

Clotfelter, C. T., Ladd, H. F., Vigdor, J. L., & Diaz, R. A. (2004). Do school accountability systems make it more difficult for low-performing schools to attract and retain high-quality teachers? *Journal of Policy Analysis and Management*, *23*(2), 251–271.

Coelli, M., & Foster, G. (2024). Unintended consequences of school accountability reforms: Public versus private schools. *Economics of Education Review*, *99*, 102523.

Deming, D. J., Cohodes, S., Jennings, J., & Jencks, C. (2016). School accountability, postsecondary attainment, and earnings. *Review of Economics and Statistics*, *98*(5), 848–862.

Dizon-Ross, R. (2020). How does school accountability affect teachers?: Evidence from New York City. *Journal of Human Resources*, *55*(1), 76–118.

Dyson, L. (2020). Walking on a tightrope: Agency and accountability in practitioner inquiry in New Zealand secondary schools. *Teaching and Teacher Education*, *93*. https://doi.org/10.1016/j.tate.2020.103075

Ehren, M. (2025). *Context, mechanisms and outcomes of accountability systems in fourteen education systems. Working paper for the Centre for Education Systems*.

Ehren, M. C. M., Altrichter, H., McNamara, G., & O'Hara, J. (2013). Impact of school inspections on improvement of schools-describing assumptions on causal mechanisms in six European countries. *Educational Assessment, Evaluation and Accountability*, *25*(1), 3–43. https://doi.org/10.1007/s11092-012-9156-4

Ehren, M. C. M., Janssens, F. J. G., Brown, M., McNamara, G., O'Hara, J., & Shevlin, P. (2017). Evaluation and decentralised governance: Examples of inspections in polycentric education systems. *Journal of Educational Change*, *18*(3), 365–383. https://doi.org/10.1007/s10833-017-9297-9

Ehren, M., Hutchinson, J., & Muñoz-Chereau, B. (2023). Place-based disparities faced by stuck schools in England: a contextual understanding of low performance and the role of inspection outcomes. *School Effectiveness and School Improvement*, *34*(4), 401–418.

Ehren, M. C. M. (2025). Context, mechanisms and outcomes of accountability systems in fourteen education systems. In *Working paper for the Centre for Education Systems*.

Eren, O., Figlio, D., Mocan, N., & Ozturk, O. (2022). *School Accountability, Long-Run Criminal Activity, and Self-Sufficiency*.

Evans, G. (2023). *Understanding the impact of school inspection: An analysis of how school leaders in Wales interpret and respond to inspection outcomes*. University of Wales Trinity Saint David (United Kingdom).

Feng, L., Figlio, D., & Sass, T. (2018). School accountability and teacher mobility. *Journal of Urban Economics*, *103*, 1–17.

Figlio, D. N., & Getzler, L. S. (2006). Accountability, ability and disability: Gaming the system? In *Improving school accountability* (Vol. 14, pp. 35–49). Emerald Group Publishing Limited.

Friedman, M. (2002). *Capitalism and freedom Milton Friedman* (I. ebrary, R. D. Friedman, & R. D. Friedman, Eds.; 40th anniversary ed.). Chicago : University of Chicago Press, 2002.

Gardezi, S. (2024). The development and status of self-evaluation as a component of school inspection and quality assurance: A comparison across four countries. *Heliyon*, *10*(20). https://doi.org/10.1016/j.heliyon.2024.e39321

Gjefsen, H. M., & Gunnes, T. (2020). How school accountability affects teacher mobility and teacher sorting. *Education Economics*, *28*(5), 455–473.

Gregg, J. J., & Lavertu, S. (2023). Test-based accountability and educational equity: Breaking through local district politics? *Economics of Education Review*, *97*, 102485. https://doi.org/https://doi.org/10.1016/j.econedurev.2023.102485

Hanushek, E. A., Saenz-Armstrong, P., & Salazar, A. (2024). *Balancing Federalism: The Impact of Decentralizing School Accountability*. http://www.nber.org/papers/w32351

Hashimoto, A. (2018). Note on the Characteristics of Japanese School Evaluation: Seven factors obstructing successful implementation. *Japanese Journal of Evaluation Studies*, *18*(2), 35–54. Hayek, F. A. von. (2001). *The road to serfdom* (2nd ed.). Routledge.

Huang, J., Yamada, H., & Zhou, Z. (2024). An analysis of the governance of the Japanese teacher license renewal system. *International Journal of Educational Research*, *128*, 102468.

Hughson, T., & Menzies, L. (2026). Polycentrism in Education Policy Making in England and Aotearoa New Zealand: Strengths as well as Weaknesses? In C. Skerritt (Ed.), *Opportunities and challenges for policy advice and policy making in education: Future possibilities*. Bloomsbury.

Hwa, Y.-Y. (2021). Contrasting approaches, comparable efficacy?: How macro-level trust influences teacher accountability in Finland and Singapore. In *Trust, Accountability and Capacity in Education System Reform*. Taylor & Francis.

Hwa, Y.-Y. (2022). 'Our system fits us': comparing teacher accountability, motivation, and sociocultural context in Finland and Singapore. *Comparative Education*, *58*(4), 542–561.

Jacob, B. A., & Levitt, S. D. (2003). *Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating*.

Jisu, R. (2020). *Supporting Professional Capacity of Teachers Through Teacher Evaluation (Doctoral dissertation)*.

Karsten, S. (1999). Neoliberal Education Reform in The Netherlands. In *Education* (Vol. 35, Issue 3).

Keane, A. (2023). Changing features of school inspection: a perspective from Wales. *Forum*, *65*(3), 81–90.

Kogan, M. (1986). *Education accountability: An analytic overview*. Routledge.

Majone, G. (1997). From the positive to the regulatory state: Causes and consequences of changes in the mode of governance. *Journal of Public Policy*, *17*(2), 139–167. https://doi.org/10.1017/s0143814x00003524

McClurg, K., Cantley, I., & Donnelly, C. (2025). The possibilities and problematics of school evaluation: post-primary teachers' perspectives on the impact of school inspections in a high stakes inspection system. *Irish Educational Studies*, *44*(3), 421–443.

McNamara, G., O'Hara, J., Brown, M., & Quinn, I. (2020). Quality assurance in Irish schools: Inspection and school self-evaluation. *Administration*, *68*(4), 161–180.

McNamara, G., Skerritt, C., O'Hara, J., O'Brien, S., & Brown, M. (2022). For improvement, accountability, or the economy? Reflecting on the purpose(s) of school self-evaluation in Ireland. *Journal of Educational Administration and History*, *54*(2), 158–173. https://doi.org/10.1080/00220620.2021.1985439

Menzies, L., & Quilter-Pinner, H. (2023). *Improvement through Empowerment: Helping our teachers and schools be the best they can be*. https://ippr-org.files.svdcdn.com/production/Downloads/improvement-through-empowerment-nov23.pdf

Moreton, H. J., Boylan, M., & Simkins, T. (2017). Headteachers who also inspect: Practitioner inspectors in England. In *School inspectors: Policy implementers, policy shapers in national policy contexts* (pp. 137–158). Springer.

Morozumi, A., & Tanaka, R. (2023). *School Accountability and Student Achievement: Neighboring Schools Matter*. Research Institute of Economy, Trade and Industry (RIETI).

Mulgan, R. (2000). 'Accountability': An Ever-Expanding Concept? *Public Administration*, *78*(3), 555–573. https://doi.org/10.1111/1467-9299.00218

Munoz Chereau, B., Hutchinson, J., & Ehren, M. (2022). *'Stuck'schools: Can below good Ofsted inspections prevent sustainable improvement?*

Nunes, L. C., Reis, A. B., & Seabra, C. (2015). The publication of school rankings: A step toward increased accountability? *Economics of Education Review*, *49*, 15–23.

O'Mara-Eves, A., Simon, A., Anand, K., Stansfield, C., Vigurs, C., & Ko, S. Y. (Jessica). (2025b). *Technical Report for Accountability Policy in Secondary Education. EPPI Centre.*

O'Mara-Eves, A., Vigurs, C., Ko, S. Y. (Jessica), & Simon, A. (2025). *Rapid systematic evidence review of school accountability policies in secondary education across fourteen jurisdictions*. EPPI Centre.

Pinto, L. E. (2016). Tensions and fissures: The politics of standardised testing and accountability in Ontario, 1995–2015. *The Curriculum Journal*, *27*(1), 95–112.

Priestley, M. (2013). Curriculum regulation in Scotland: A wolf in sheep's clothing is still a wolf. *European Conference on Curriculum Studies*.

Proudfoot, K. (2025). Creeping anglicisation: school inspection in Scotland and Ireland. *Comparative Education*, 1–18.

Rezai-Rashti, G. M., & Segeren, A. (2023). The game of accountability: perspectives of urban school leaders on standardized testing in Ontario and British Columbia, Canada. *International Journal of Leadership in Education*, *26*(2), 260–277.

Sahlberg, P. (2016). The global educational reform movement and its impact on schooling. *The Handbook of Global Education Policy*, 128–144.

Şen, A. Y., & Maya, İ. (2022). School quality assurance applications for basic education schools. *Education Quarterly Reviews*, *5*.

Smith, A. [1776]. (2005). *An inquiry into the nature and causes of the wealth of nations: a selected edition*. AN ELECTRONIC CLASSICS SERIES PUBLICATION. https://www.rrojasdatabank.info/Wealth-Nations.pdf

Torres, R. (2021). *Does test-based school accountability have an impact on student achievement and equity in education?: A panel approach using PISA*.

West, A., Mattei, P., & Roberts, J. (2011). Accountability and Sanctions in English Schools. *British Journal of Educational Studies*, *59*(1), 41–62. https://doi.org/10.1080/00071005.2010.529416

World Bank. (2008). *What is school-based management?* World Bank. http://documents.worldbank.org/curated/en/113901468140944134

CMO Configurations

We conclude by summarising a series of hypothetical 'context – mechanism – outcome configurations' which may help to explain how accountability generates its effects, under different conditions. These are not uniquely associated with a particular approach and instead can arise in different systems depending on conditions and context.[193]

1. Pressure to conform to norms and expectations (coercive institutional isomorphism):

Coercive pressures from accountability systems can lead organisations within the same sector to copy one another and become similar in shape, structure or form.

The effects of this mechanism can be desired, undesired or ambiguous. These effects might include:

- More targeted improvement
- Improved teaching practices
- Increases in student performance
- Schools increasingly meeting expected standards
- Narrowing of the curriculum and discouraging new teaching methods
- Centralisation of decision-making in schools
- Heightened anxiety and frustration (among teachers and/or pupils)

This mechanism has been linked to a range of accountability practices. These include risk-based inspections, test-based accountability and league tables, school inspections and published reports, performance management systems, assessments, (mandated) school self-evaluations and combinations of these. The threat of reduced autonomy may prompt compliance (ironically – a desire not to lose autonomy prompts a partial yielding of it).

The extent to which different effects prevail may be linked to conditions.
These include schools' capacity to respond, how well they understand the expectations, and whether schools risk losing autonomy in response to poor performance. Societal expectations can be another mediating factor, for example the degree to which the public considers accountability measures, and the expected standards to be legitimate.

How school leaders respond can also shape effects. Some Heads act as a 'buffer' – protecting their staff; others explain the outcomes of accountability to their staff and include them in responding and implementing changes, rather than simply fulfilling administrative requirements through a compliance-focused orientation.

[193] Full details available in Ehren (2025)

2. Convergence around professional norms (normative institutional isomorphism):

When members of an occupation collectively define the conditions and methods of their work, this can drive non-coercive 'normative' convergence. This might be driven by professional training, certification, shared standards and interaction (for example through professional networks).

Normative convergence might also have both desired and undesired effects. These effects might include:

- Shifts in teachers' beliefs and practices
- Changes in school leadership, policy and practice
- Professional growth among teachers
- Improved student outcomes
- Stress, high workload and gaming.

This mechanism has been linked to various accountability practices such as:

- Self-evaluation and development planning
- Teacher evaluations
- Training educators to serve as peer inspectors on inspection teams and share their learning more widely
- The use of inspection standards to inform wider decision-making and legitimate the sharing of good practice.

The prevalence of positive versus more negative effects might depend on schools' degree of autonomy and capacity. This is because autonomy and capacity might provide teachers and leaders with more opportunities to engage with standards meaningfully.

As with more coercive forms of convergence, principals can react differently to accountability systems in different conditions. Some pursue superficial administrative compliance, whereas others work with staff to implement meaningful change. Teacher training and related support policies may help professionals to respond more positively. In contrast, a lack of clarity about expectations may trigger 'hyper-enactment' more akin to coercive convergence.

School boards and other stakeholders can sometimes play a role in promoting normative convergence. Self-evaluation might provide an opportunity to discuss standards and create buy-in. This is perhaps easier when there is sufficient autonomy and capacity at a school level. Alignment between accountability standards and existing norms or ethos – combined with early feedback – might also support professional-norm driven convergence.

3. Adaptation due to competitive pressures (competitive isomorphism):

Educational institutions may abandon or adopt practices in order to survive in the face of competitive pressure. This mechanism is largely driven by market pressures.

Competition driven adaptation can give rise to:

- 'Teaching to the test' – including focusing on assessed aspects of the curriculum and assessed subjects. This may or may not be desirable;
- Pupil anxiety;
- Increased applications to high performing schools.

This mechanism tends to be linked to market based accountability, where parents use exam results and school rankings as part of school choice. It might be more likely to arise when parents consider publicly available information to represent relevant factors, and when there is an adequate local supply of provision to enable choice. Declining student numbers might exacerbate pressures, as might pressure from the school community due to interactions with house values.

Where teachers themselves are the actors in an accountability system (particularly under vertical teacher accountability), competition can happen between individuals rather than institutions. This may give rise to the same mechanisms and similar effects, particularly where consequences extend to pay and career progression.

4. Organisational learning:

Educational institutions can innovate and learn through collaboration and reflection. Accountability can contribute to this and appears to be a relatively widespread mechanism associated with numerous forms of accountability.

Organisational learning might happen when accountability prompts staff to:

- Reflect on the quality of their work, and to use assessment data and inspection feedback for improvement;
- Evaluate and improve teaching and learning through collective decision-making;
- Reconsider existing views of their school's performance in response to a fail grade and provide momentum for change.

Certain conditions may help to ensure schools respond to accountability through organisational learning. These include:

- Limited immediate pressure;
- Actionable feedback;
- School or district inspectors supporting learning through context-sensitive professional conversations;

- Allocating additional resources to support improvement may also be conducive to organisational learning;
- Heads working as peer inspectors;
- A tradition of quality assurance;
- A high degree of trust and collaboration.

Organisational learning may be undermined when there is a lack of emphasis on developmental planning and self-evaluation, and where communication between teachers and principals is characterised by stigmatisation and strategic behaviour. This dynamic may be further exacerbated by punitive cultures; a perception that accountability is an intrusion into teachers' professional fiefdom; a view that teachers cannot or should not be held responsible for pupil achievement; and poor assessment literacy.

5. Self-determination:
Individuals are likely to be more (intrinsically) motivated when their basic psychological needs for autonomy, competence and relatedness are met. When accountability systems meet these needs, this may lead to:

- Increased individual motivation;
- Better quality teaching;
- Strengthened professional ethics.

When accountability systems undermine individuals' psychological needs this may result in:

- Increased stress and anxiety;
- Performativity and demotivation.

Hierarchical, less respectful approaches, and frequently shifting targets might undermine teachers' and leaders' basic needs.

Appendix 1: Methods

This report is based on four components and further details of the methods used are available in those outputs.

Research Questions

Six main research questions guided this study, along with sub questions which were refined through policy stakeholder workshops.

RQ1. What is the purpose of accountability?

a. What are the stated aims of accountability?
b. Are there particular values (for example 'fairness' / 'excellence' / 'meritocracy') associated with accountability?
c. What if any shifts in priorities can be found in official communications and documentation?

RQ2. What is the structure of accountability?

a. At what level is accountability directed? (Who is held accountable by whom? For what? In what way? And with what consequences?)
b. What are the most common types of accountability?
c. What mechanisms are used or assumed to link accountability to improvement and is there a plan for how these different mechanisms are intended to interact or relate to each other?
d. Is there a distinct regulatory/school authorisation function – what is the relationship between the reporting function and regulatory function?

RQ3: How is accountability policy made and introduced?

a. What is the process for managing change in accountability policy and which, if any, bodies exist to manage this process?
b. How was the most recent accountability reform designed and implemented, and who was involved?

RQ4: How is accountability policy evaluated?

a. Are there regular or ad-hoc assessments of the accountability system's effectiveness?
b. Are these evaluations done internally or by an independent partner, outside of government?
c. What, if any, success metrics are used to assess the effectiveness of the accountability system?
d. How if at all, are unintended consequences (e.g. on leadership and teacher agency and motivation) monitored?

e. To what extent is there evidence that policy evaluation has led to changes in policy?

RQ5: How does context shape accountability policy and reform?

a. How have historic and cultural factors, including the country's 'education policy orientation' influenced its accountability system?
b. How has context (including political circumstances – eg electoral cycle) shaped (and how is context shaping) recent (and upcoming) accountability reforms?
c. What level of coherence is there with other elements of the education system?

RQ6: What is the evidence about the effects of the accountability system?

a. What does research from each jurisdiction say about the intended and unintended impact of the accountability system on outcomes (including different types of outcomes and different groups)
b. Looking across jurisdictions, what themes emerge from the available research evidence relating to the impact of accountability systems?
c. What does research from each jurisdiction say about key stakeholders' views and experiences of each country's accountability system.
d. Looking across jurisdictions, what themes emerge from the available research evidence relating to stakeholders' views and experiences of accountability systems?
e. Looking across jurisdictions, what does the research evidence and stakeholder interviews tell us about the context, mechanisms, and outcomes of accountability systems?
f. What do causal studies suggest are the effects of accountability systems (in general)?
g. To what extent were each country's latest accountability reforms implemented as planned?

Case Selection

The project focused on fourteen jurisdictions.: England, Scotland, Wales, Northern Ireland, the Republic of Ireland, Ontario, New Zealand, Singapore, Japan, Estonia, Finland, Poland, The Netherlands and France.

Jurisdictions were chosen based on diverse case selection nested within a similar domain.[194] More specifically, jurisdictions were:

[194] Seawright and Gerring (2008)

1. High or upper-middle income democracies which achieve near or above average results in PISA.
2. A combination of larger and smaller systems with different political and cultural histories.

Additionally, jurisdictions that are of interest to CES' target audience of policy makers – primarily in the UK – were particularly prioritised. The final country selection was agreed following consultation with policy stakeholders from across all four UK Home Nations.

Research Strategy

The first four research questions were largely descriptive and were addressed through a review of up to date policy documentation, combined with validation involving in-country experts. RQs 5 and 6 were more analytic and were approached through systematic reviews of different forms of secondary evidence, alongside validation with in-country experts (see below).

A realist lens was applied to the study[195] including:

- **A theory-informed approach:** This involved building on the foundations laid by previous research and gradually refining these theories over the course of the study.
- **A focus on generative causality:** This involved asking how certain patterns of events might be generated and what mechanisms underlie this'.
- **An emphasis on context:** This involved paying attention to "the circumstances in which a particular causal trajectory is more likely and less likely to follow" and identifying the boundary conditions within which a pattern occurs.[196]
- **Theory informed in-case and cross-case comparison:**[197] This involved theory-informed 'shuttling'[198] between empirical case-by-case and cross-case data.

The study was specifically designed to provide actionable evidence for policy makers. The aim was not to reach a judgement on 'what works' but to identify different alternatives, the extent to which these might 'travel' across contexts,and what effects might be associated with different approaches.

There were four stages to the work.

1. **Study scoping and theoretical positioning.**
    - Development of a 'Concept Note' (Appendix 2)[199] outlining the studies theoretical background and establishing the study's parameters.

[195] Pawson (2024)
[196] Pawson (2024) p243
[197] George and Bennett (2005)
[198] Atkinson et al (2003)
[199] Ehren (2024)

- Two workshops with policy stakeholders across the four UK home nations to refine questions, cases, and priorities.

2. **In-case analysis**
    - Development and validation of detailed country reports, structured around RQs and subquestions (primarily Work Package 1).

3. **Cross-case analysis**
    - Development of cross-country comparative tables
    - Evidence reviews (Work Packages 1 and 2)

4. **Theory development and refinement**
    - Development of Context-Mechanism-Outcome Configurations through a Realist Inquiry (Work package 3)
    - Synthesis and report writing (Work Package 4)

The work packages

Work Package 1: Led by Alison O'Mara-Eves and Antonia Simon
The research team also included: Kusha Anand, Carol Vigurs and Jessica Ko Sum Yue.

There were two aspects to this strand of work.

i. **A pragmatic review of policy documentation via searches of government and educational body websites, with expert validation.**[200] This aspect of the study focused on the descriptive research questions (RQs 1-4). Information from policy documentation was extracted and summarised within each of the fourteen country reports. Each jurisdiction report followed a common structure based on the research questions, informed by the 'Concept Note'. In-country experts were consulted to review the draft reports documents. They checked for major inaccuracies, suggested information for gaps, and suggested corrections and additions where appropriate. Wherever possible, this 'sense check' validation was carried out in consultation with two policy makers per jurisdiction with experience in school accountability (however this was not possible in all cases). Once all fourteen country profiles had been completed, a series of tables were created to provide descriptive comparisons of the jurisdictions, using a 'machine in the loop', AI assisted approach.[201]

ii. **A rapid systematic review of the research evidence on RQ 5 and 6:**[202] For the most part, this followed the processes of a full systematic review. It was 'rapid' in four ways: the use of machine learning to sift out studies less likely to be relevant to

[200] O'Mara-Eves et al. (2025b)
[201] Hong et al., 2022
[202] O'Mara-Eves et al (2025)

the review topic; the lack of citation chasing of the included studies; the use of a 'light-touch' critical appraisal tool; the use of AI to support aspects of data extraction and identifying themes across studies. 107 unique studies were included. Experts were also consulted at this stage to check important studies were not being omitted. Data for each jurisdiction was thematically summarised and presented within each country report. Additionally, a cross-cutting thematic analysis was carried out comparing themes running across the jurisdictions. These were presented in a separate thematic report.

Work package: Led by Simon Burgess[203]

This strand of work focused on the effects of accountability systems. It involved a highly targeted approach to identifying quantitative studies that describe (potentially generalisable) causal effects linked to accountability systems- typically through experimental or quasi experimental approaches. Because of the limited pool of robust studies meeting this criteria, studies in jurisdiction beyond CES 14 focus jurisdictions were included. The review's approach was to conduct purposive searching to update Burgess and Greave's 2021 study of accountability effects,[204] and included studies identified as part of WP1 that met WP2's narrower criteria.

Work package 3: Led by Melanie Ehren[205]

WP3 brought together RQ5 and 6 by identifying hypothetical 'context-mechanism-outcome' (CMO) configurations through a realist inquiry. This involved

- Reading each jurisdiction profile in full to understand the accountability system, its context and purpose and its development over time.
- Reviewing the summaries of each research study that had been included in the profile, and describing:
    - the elements of the accountability system included in the research,
    - their outcomes,
    - how (elements of accountability) generate certain outcomes (mechanisms)
    - under what conditions (the context).
- These descriptions were shared with in-country experts who provided feedback. They were then integrated into one or more CMO configurations, linking these to elements of accountability where possible.

For each of the CMO configurations, an evidence statement was written, describing the underlying sources and how these were interpreted to construct the CMO. The rigour and relevance of underlying sources was described using established classifications.[206]

[203] Burgess (2025)
[204] Burgess and Greaves 2021
[205] Ehren (2025)
[206] Eddy-Spicer et al (2016), p.36; Pawson, (2006), p.22

Work Package 4 (this report): Led by Loic Menzies

This work package involved further cross-case analysis and synthesising findings from across the three preceding work packages. Outputs from the work packages were read in detail and additional comparative tables were generated, returning to original sources and seeking additional detail where necessary.

- **RQs 1-4:** Cross-case findings were summarised using a structure that was informed by the research sub-questions, the 'Concept Note', and additional theoretical insights from the literature. Diagramatic representations – such as the 'accountability webs' were created wherever possible and this facilitated the description of recurring patterns, the development of 'ideal types' and typologies.[207]
- **RQ5:** Findings on accountability contexts from EPPI's thematic report were summarised.
- **RQ6:** The WP2 and the cross-jurisdictional findings from WP3 were reread and all causal claims about effects extracted and entered into a table.
    - the accountability element – eg inspection;
    - the approach – eg 'publication of performance tables by government';
    - the corresponding accountability approach
    - country;
    - any contextual considerations (eg high trust, high availability of school choice);
    - potential mechanisms (e.g. teacher turnover results in a replacement of staff);
    - effects

  Effects were grouped into types using tags (e.g. 'workforce', 'teacher/school practices). The EPPI thematic report on RQ6 was reviewed alongside this, to ensure no major findings had been missed (if so they were added to the table). Findings were then written up thematically by effect type, returning to original sources and country profiles wherever necessary.

[207] George and Bennett 2005

Appendix 2: Concept note

Conceptualising 'Accountability' for the comparative review
Melanie Ehren, 2024

Summary

School accountability involves educational institutions having an obligation to report on, explain or justify what they have done to one or more sources of authority, drawing on various types of information. This may be done through formal or informal means, but some form of judgement will normally be reached, usually leading to a consequence.

'Information' can come from a range of quality measures, including (but not limited to) school inspections and standardised tests. Various organisations and individuals may be involved in reaching decisions based on that information, including (but not limited to) school managers, inspectorates, departments of education, and the wider stakeholder community (including parents choosing schools).

Accountability is 'a social relation between an actor and a forum, where the actor has the obligation to explain and justify conduct, where information is provided and debated, a forum passes judgement, and the actor faces consequences'.

Within these accountability relations there can be different types of quality measures (e.g. school inspections, standardised tests) that provide information on distinct aspects of quality and performance of educators and school organisations, provide information for decision-making to different stakeholders (the forum), and have varying intended and unintended consequences for involved actors. Specific combinations of who is held accountable to whom, and the measures and consequences within the accountability relationship, can be described as ideal types of school accountability systems, such as bureaucratic, marketised, or performance-based accountability.

Background

Accountability is a broad concept that can have different meanings, depending on the field of study or the specific scope, functions, and goals chosen in a particular jurisdiction. This working paper aims to conceptualise our understanding of accountability with the purpose of informing the various studies that are part of the comparative review. We start with the following organising principles, which were agreed with the team on 2 October 2024 and result from the various stakeholder meetings:

- Our search terms and protocol need to be sufficiently open to find whatever the jurisdictions consider to be 'accountability' so that we can note divergence from our working definition.
- However, we should still discuss and agree on broad parameters, based on the team's existing knowledge, including relevant typologies. We can then use this to provide working definitions that will be refined in light of our data.
- Because our primary interest is in the country's approach to accountability policy, our focus is on jurisdiction-level accountability for schools (rather than regions, municipalities, and teacher accountability).
- However, we may find evidence that, for example, there is very little school accountability but there is a lot of accountability at a teacher level. We would need to note this evidence as a contextual consideration.

Defining school accountability

Formal versus informal

'In its most general sense, accountability means giving a justification of what one has done'[208]. Such justifications can be given in an informal and nonroutinized manner, or through more formalized systems. The informal manner involves simply giving an account, taking responsibility for explaining one's actions in a form of social interaction, while the latter implies being called to give an account by an authority who enforces responsibility on another through more standardized and prescribed systems[209].

Parameter 1: a focus on formal accountability

Given the aim of the centre is to inform policy making, we will focus on formal types of school accountability that typically consist of codified standards and measures with certain consequences for reaching or for not reaching the standards. 'Formal' refers to the measures used and how information is shared. Consequences, however, can be both formal (and intended), such as when inspection outcomes are used to allocate support and resources by government agencies, as well as informal (and sometimes unintended), such as when stakeholders act on outcomes in ways unexpected in the original design of the system.

[208] Hoffer (2000) p529
[209] Mulgan, R. (2000)

Parameter 1: a focus on formal accountability
Given the aim of the centre to inform policy, we will focus on formal types of school accountability that typically consist of codified standards and measures with certain consequences for reaching or for not reaching the standards. 'Formal' refers to the measures used and how information is shared. Consequences however can be both formal (and intended), such as when inspection outcomes are used to allocate support and resources by government agencies, as well as informal (and sometimes unintended), such as when stakeholders act on outcomes in ways unexpected in the original design of the system.

Accountability is relational: actors and a forum

Accountability implies a relationship between those who seek answers and those who are answerable, those who may impose sanctions and those who anticipate them, according to Brady.[210] This type of relationship is captured in Bovens' commonly used **definition of accountability** as:

'a social relation between an actor and a forum, where the actor has the obligation to explain and justify conduct, where information is provided and debated, a forum passes judgement, and the actor faces consequences'. [211]

The actor can be an individual, but will often be an organisation; the forum can be a specific person (professional peer, journalist, minister, superior), an agency (inspection agency, parliament, audit office, the court), or a virtual entity (the general public or specific advocacy group).

In our context, actors are individuals and organisations in the education system, such as a teacher, a headteacher or a school. The significant other, called the 'accountability forum' by Bovens, can be the general public, central government or the parents of the children in the school.

De Grauwe[212] and Hooge et al.[213] distinguish the following four fora as the most common in monitoring systems in education:

1. the education administration, represented by the ministry (public/state control model);
2. the teachers themselves (in a peer review model; also referred to as the professional accountability model);
3. parents in close relationship with the school (the partnership model);

[210] Brady (2021). p26
[211] Bovens (2007) p9
[212] De Grauwe (2007)
[213] Hooge et al. (2012) p8

4. the general public by means of parental choice and competition between schools (the free market model or market accountability).

Hooge et al.[214] describe the first accountability relationship between schools and the education administration as vertical accountability: 'top-down and hierarchical, enforcing compliance with laws and regulation and/or holding schools accountable for the quality of education they provide'. The second and third type of relation is horizontal and presupposes non-hierarchical relationships; they are directed at how schools and teachers conduct their profession, and/or at how schools and teachers provide multiple stakeholders with insight into their educational processes, decision making, implementation, and results.

Parameter 2: actor and forum in horizontal and vertical accountability relations
For our review, we define 'actor' in the accountability relationship as the individual school organisation and school groupings (e.g. Multi-Academy Trust). We are interested in both their horizontal accountability (e.g. to other schools in a grouping) and their upward (vertical) accountability towards a government agency at the local, district, provincial and/or national level (forum), as well as their accountability to the general public – and, more specifically, to pupils and parents – through, for example, school performance tables or forms of community accountability.

Quality concept and measures

A further distinction can be made in *what* needs to be explained and *how* that is measured and judged, or the quality concept underpinning the accountability and specific measures used. De Grauwe[215] refers to the focus of monitoring systems and these can include:

- School inputs, and norms and regulations.
- Instructional and school organisational processes and characteristics (e.g. teaching and learning, school climate, leadership and decision-making).
- School output: both short and long-term results, such as student achievement, graduation rates and students' future schooling and employment career.

The different quality concepts can overlap but require different types of measures and judgements. A focus on school input, norms and regulations tend to come with management information systems, audits or inspections that check for compliance. Measures of process and school organisation often make use of multiple sources, such as quantitative data from parent and student surveys; qualitative data from interviews and focus groups; and school documents and lesson observations. A measurement of process also requires data collection from different actors, inside and outside the school organisation according to Hooge et al.[216] Process measures can also include peer reviews,

[214] Hooge et al. (2012) p8
[215] De Grauwe (2007a) p18
[216] Hooge et al (2012)

self-assessments, or the involvement of a more diverse set of evaluators (e.g. experts, critical friends, parents). Accountability systems that focus on school output, such as performance-based accountability systems, generally use standardised student assessments with underlying models to aggregate student results at the school level and correct for students' socio-economic backgrounds.

Parameter 3: quality concept and measures

Our review will include all types of quality concepts and related measures – school inputs, processes, and outputs – and the types of measures used to hold schools accountable, such as management information systems, school inspections, standardised tests, etc.

Purpose

Accountability systems vary in their intended purpose. De Grauwe[217] describes three different, yet complementary roles of school inspections that could also apply to other types of evaluation. These are 'control and evaluate', 'improvement and support' (including giving advice), and 'liaison'. The latter purpose or role is generally attributed to Inspectorates of education as these are typically positioned between the top of the education system where norms and rules are set, and the schools who need to implement these. In their role as go-between agents, they can have a task to inform schools of decisions taken by the centre, and to inform the centre of the realities at school level. Ladd[218] similarly talks about quality assurance and the goal of accountability systems to assure that all schools are as high quality (defined in terms of their internal processes) as possible given the resources available to them. A separate goal, in her view, is improved student outcomes, while a third is distributional equity – a term that refers to how some groups of children fare relative to others. This objective might be interpreted in terms of either outcomes or internal processes and practices.

Parameter 4: purpose of accountability

We are interested in all the purposes of school accountability.

Consequences (formal and informal), including use of accountability information and outcomes

Consequences of accountability typically include a set of sanctions and interventions (monitoring/support) when failing to meet a predefined target, but may also involve praise or special entitlements for excellent performance. Wößmann et al.[219] categorise consequences as positive (rewards) or negative (sanctions), and they may be implicit

[217] De Grauwe (2007b)
[218] Ladd and Thompson (2012)
[219] Wöbmann et al. (2007)

(e.g. the respect of peers) or explicit (e.g. cash bonuses). Consequences are often discussed in terms of being low or high stakes, with Braithwaite's intervention pyramid[220] ranking them from low to high. However, even when accountability systems, by design, are thought to be low stakes, educators tend to view them as high stakes, particularly when there is reputational damage involved and schools are functioning in a competitive market.[221] Consequences can include the immediate sanctions from failing to meet an accountability target, but are often broader when other actors in the system use accountability outcomes in their decision-making. This is particularly relevant in systems that provide accountability information to the general public or stakeholders for use in school choice or to target school improvement. We therefore include the intended sharing of accountability information with intended users in the consequences from accountability.

Parameter 5: consequences of accountability
We include all types of intended consequences (rewards, sanctions, support and additional monitoring) in our definition of accountability, including the intended sharing of accountability information.

Types of school accountability

We can distinguish different types of accountability systems on the basis of specific configurations of *who* is accountable to *whom, for what*, *how,* and with what type of *consequences*. Various authors present these as ideal-types, but they tend to be much more diffuse in reality with various combinations of systems. A brief summary is presented below by the type of relation (vertical/horizontal) that is prioritised, the quality concept and focus of the accountability relationship, and the level of consequences involved.

Vertical/horizontal relations
Hooge et al.[222] present four typologies that fit with their distinction in vertical and horizontal accountability relations and include different actors as well as quality concepts:

Vertical	*Regulatory school accountability*: compliance with laws and regulations; focuses on inputs and processes within the school.
	School performance accountability: periodic school evaluations.
Horizontal	*Professional school accountability*: professional standards for teachers and other educational staff.
	Multiple school accountability: involving students, parents and other stakeholders in formulating strategies, decision-making, and evaluation.

[220] Lodge (2015)
[221] See Munoz et al. (2022); Jones et al (2017)
[222] Hooge et al. (2012) p.9

These four models, to some extent overlap with De Grauwe's[223] presentation of four types by who is responsible for the quality monitoring and improvement of schools:

1. The public or state control model: where employers or political power holders are in charge of monitoring. The main characteristic is that of a formal bureaucratic hierarchy: teachers are controlled by school headteachers, who are controlled by district officers, who are controlled by central ministries that in turn are directed by elected representatives.
2. The professional accountability model: the professional community, e.g. the teaching force, is in charge of monitoring through ethical codes of conduct which are monitored through self-evaluation by teachers, peer review and/or wider systems for certification.
3. The partnership model: based on a partnership between the parents and/or students and teachers/the school where the accountability relationship involves consensus on objectives, an exchange concerning methods, and a discussion about results obtained. The partnership model assumes parity between the providers of education and the clients and where internal decisions about school functioning are shared. The main monitoring device will be self-assessment through heavy involvement of parents along with shared decision-making.
   Hooge et al.[224] call this model 'participative democracy' which includes the establishment of councils and committees that actively involve teachers, parents and pupils, alongside other members of the community, in shared decision-making. In the same vein, we find references to social, or community accountability.[225]
4. The free market model: emphasises the control of the individual consumer, such as through school choice, by providing parents with vouchers and allowing them to buy the preferred schooling for their children. Hooge et al. refer to a market mechanism which is achieved by enhancing parental choice and encouraging school competition.[226] Policy arrangements to abolish catchment areas, create voucher programmes and set up charter schools are also commonly used mechanisms for market-based accountability.

Focus/quality concept

De Grauwe, Eddy-Spicer et al. and Ladd describes three types of systems by the focus or type of quality concept that is monitored:[227]

1. Compliance monitoring (also referred to as 'regulatory school accountability'): the emphasis is on school inputs (number of text books per pupil, teacher qualifications, number of pupils per class, etc.). It has been called compliance monitoring as its first

[223] De Grauwe (2007a) p.14–16
[224] Hooge et al. 2012
[225] Boelen and Woollard, B. (2009); Westhorp et al. (2014).
[226] Hooge et al. (2012)
[227] De Grauwe (2007a) p12; Eddy-Spicer et al. (2016); Ladd (2012)

goal is to make sure that schools comply with predetermined norms fixed by law and administrative rules and regulations. Compliance monitoring is the oldest bureaucratic type of monitoring: checking that rules and regulations are respected. The classic inspectorate system, combined with several forms of administrative self-reporting by schools (filling out forms), is the main device on which this type of monitoring relies.

2. Diagnostic monitoring: the goal of this type of monitoring is to ensure that pupils learn what they are supposed to learn. The focus is on the instructional process, on what happens in the classroom. We can extend the focus to the school organisation (leadership, culture etc) and the standards often included in school inspection frameworks.
3. Performance monitoring: the emphasis of this type of monitoring is on school results. Its goal is often to stimulate competition between schools in order to promote academic achievement. The most common monitoring devices used are the regular measurement of learner achievement by standardised tests and examinations, combined with the publication of league tables and systematic (external) auditing of schools. Performance monitoring underpins most of the accountability in education in the United Systems and is frequently described as high stakes testing, test-based accountability, standards-based reform or school accountability.[228] The latter term is somewhat confusing as 'school accountability' can include other types of measures and the Every Student Succeeds Act (ESSA, introduced in 2015) in the US particularly promotes the use of multiple and broader measures of success[229]. Ladd describes test-based accountability as a system where 'policy makers hold schools accountable for the performance of their students as measured by test scores or other outcome measures such as graduation rates'.[230]

Low/high stakes systems

Accountability systems are often also categorized as low or high stakes, referring to the consequences of school accountability and how these are perceived by actors in the accountability relationship.

[228] Bae (2018); Darling-Hammond, L. (2004); Figlio and Loeb (2011) Kane and Staiger (2002); Linn, R. L. (2008).
[229] Darling-Hammond et al. (2016)
[230] Ladd (2012) p2

Complete Reference List

NB: References for policy documents and additional sources can be found in the full and underlying reports and tables in the CES library on the CES website.

Altrichter, H., & Kemethofer, D. (2015). Does accountability pressure through school inspections promote school improvement? *School Effectiveness and School Improvement*, *26*(1), 32–56. https://doi.org/10.1080/09243453.2014.927369

Anderson, P. M., Butcher, K. F., & Schanzenbach, D. W. (2017). Adequate (or adipose?) yearly progress: assessing the effect of "No Child Left Behind" on children's obesity. *Education Finance and Policy*, *12*(1), 54–76.

Andrabi, T., Das, J., & Khwaja, A. I. (2017). Report cards: The impact of providing school and child test scores on educational markets. *American Economic Review*, *107*(6), 1535–1563.

Atchison, Drew, Le Floch, Kerstin Carlson, Özek, Umut, Luu, Justin, Hurlburt, Steven, & Boyle, Andrea. (2025). The Effects of School Accountability Ratings in California Under ESSA on Student Graduation and College and Career Readiness. *AERA Open*, *11*, 23328584251349184. https://doi.org/10.1177/23328584251349186

Atkinson, P., Coffey, A., & Delamont, S. (2003). *Key themes in qualitative research: Continuities and changes*. Rowman Altamira.

Bae, S. (2018). Redesigning systems of school accountability: A multiple measures approach to accountability and support. *Education Policy Analysis Archives*, *26*, 8.

Ball, S. J. (2009). Academies in context: Politics, business and philanthropy and heterarchical governance. *Management in Education*, *23*(3), 100–103.

Ball, S. J., Bowe, R., & Gewirtz, S. (1996). School choice, social class and distinction: the realization of social advantage in education. *Journal of Education Policy*, *11*(1), 89–112. https://doi.org/10.1080/0268093960110105

Ball, S. J., & Junemann, C. (2012). *Networks, new governance and education*. Policy Press Bristol.

Bassok, D., Markowitz, A. J., Bellows, L., & Sadowski, K. (2021). New evidence on teacher turnover in early childhood. *Educational Evaluation and Policy Analysis*, *43*(1), 172–180.

Baumgartner, F. R., Jones, B. D., & Mortensen, P. B. (2023). Punctuated equilibrium theory: Explaining stability and change in public policymaking. *Theories of the Policy Process*, 65–99.

Bergbauer, A. B., Hanushek, E. A., & Woessmann, L. (2024). Testing. *Journal of Human Resources*, *59*(2), 349–388.

Bertoni, M., Brunello, G., & Rocco, L. (2013). When the cat is near, the mice won't play: The effect of external examiners in Italian schools. *Journal of Public Economics*, *104*, 65–77.

Boelen, C., & Woollard, B. (2009). Social accountability and accreditation: a new frontier for educational institutions. *Medical Education*, *43*(9), 887–894.

Bokhari, F. A. S., & Schneider, H. (2011). School accountability laws and the consumption of psychostimulants. *Journal of Health Economics*, *30*(2), 355–372.

Bovens, M. (2007a). Analysing and assessing accountability: A conceptual framework1. *European Law Journal*, *13*(4), 447–468. https://doi.org/10.1111/j.1468-0386.2007.00378.x

Bovens, M. (2007b). Analysing and assessing accountability: A conceptual framework1. *European Law Journal*, *13*(4), 447–468. https://doi.org/10.1111/j.1468-0386.2007.00378.x

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2008). The impact of assessment and accountability on teacher recruitment and retention: Are there unintended consequences? *Public Finance Review*, *36*(1), 88–111.

Brady, A. M. (2021). Response and Responsibility: Rethinking Accountability in Education. *Journal of Philosophy of Education*, *55*(1), 25–40. https://doi.org/10.1111/1467-9752.12501

Brown, M., McNamara, G., O'Hara, J., & Shevlin, P. (2020). Polycentric inspection: A catalyst for sparking positive interactions in educational networks. *Evaluation*, *26*(1), 76–97.

Burgess, S. (2025). *A review of quantitative causal research evidence on school accountability: updating the evidence*.

Burgess, S., & Greaves, E. (2021). School choice and accountability. In *Oxford Research Encyclopedia of Economics and Finance*.

Burgess, S., Wilson, D., & Worth, J. (2013). A natural experiment in school accountability: The impact of school performance information on pupil progress. *Journal of Public Economics*, *106*, 57–67.

Burns, T., Köster, F., & Fuster, M. (2016). *Educational research and innovation education governance in action lessons from case studies: Lessons from case studies*. OECD Publishing.

Camargo, B., Camelo, R., Firpo, S., & Ponczek, V. (2018). Information, market incentives, and student performance: Evidence from a regression discontinuity design in Brazil. *Journal of Human Resources*, *53*(2), 414–444.

Centre for Governance and Scrutiny. (2010). *Accountability works! How we can make it better*.

Centre for Governance and Scrutiny. https://www.cfgs.org.uk/wp-content/uploads/Accountability-works.pdf

Chiang, H. (2009). How accountability pressure on failing schools affects student achievement. *Journal of Public Economics*, *93*(9–10), 1045–1057.

Cilliers, J., Mbiti, I. M., & Zeitlin, A. (2021). Can Public Rankings Improve School Performance?: Evidence from a Nationwide Reform in Tanzania. *Journal of Human Resources*, *56*(3), 655–685.

Clapham, A. (2015). Post-fabrication and putting on a show: examining the impact of short notice inspections. *British Educational Research Journal*, *41*(4), 613–628.

Clotfelter, C. T., Ladd, H. F., Vigdor, J. L., & Diaz, R. A. (2004). Do school accountability systems make it more difficult for low-performing schools to attract and retain high-quality teachers? *Journal of Policy Analysis and Management*, *23*(2), 251–271.

Coelli, M., & Foster, G. (2024). Unintended consequences of school accountability reforms: Public versus private schools. *Economics of Education Review*, *99*, 102523.

Committee on Standards in Public Life. (1995). *The Seven Principles of Public Life*. GOV.UK. https://www.gov.uk/government/publications/the-7-principles-of-public-life/the-7-principles-of-public-life--2

Dandala, S. (2019). Human resource policy and teacher appraisal in Ontario in the era of professional accountability. *Management in Education*, *33*(1), 5–10.

Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, *106*(6), 1047–1085.

De Grauwe, A. (2007a). *Module 7. Alternative models in reforming school supervision*. UNESCO International Institute for Educational Planning. http://www.iiep.unesco.org/fileadmin/user_upload/Cap_Dev_Training/Training_Materials/Supervision/SUP_Mod7.pdf

De Grauwe, A. (2007b). Transforming School Supervision into a Tool for Quality Improvement. *Education*, *53*(5), 709–714. https://doi.org/10.1007/sl

Dee, T. S. (2020). Learning from the Past: School Accountability before ESSA. *Stanford, CA: The Hoover Education Success Initiative. Retrieved January*, *7*, 2021.

Deming, D. J., Cohodes, S., Jennings, J., & Jencks, C. (2016). School accountability, postsecondary attainment, and earnings. *Review of Economics and Statistics*, *98*(5), 848–862.

Department for Education (Ireland). (2022). *School Self-Evaluation: Next Steps* . https://www.pdst.ie/sites/default/files/SSE_next_steps.pd

Direction de l'évaluation, de la prospective et de la performance. (2025, September). *Indicateurs de valeur ajoutée des lycées*. Direction de l'évaluation, de La Prospective et de La Performance (DEPP). https://catalogue.depp.education.fr/index.php/catalog/274

Dizon-Ross, R. (2020). How does school accountability affect teachers?: Evidence from New York City. *Journal of Human Resources*, *55*(1), 76–118.

Dubnick, M. (2005). Accountability and the promise of performance: In search of the mechanisms. *Public Performance & Management Review*, *28*(3), 376–417.

Dubnick, M. J. (2002). Seeking salvation for accountability. *Annual Meeting of the American Political Science Association*, *29*, 7–9.

Dyson, L. (2020). Walking on a tightrope: Agency and accountability in practitioner inquiry in New Zealand secondary schools. *Teaching and Teacher Education*, *93*. https://doi.org/10.1016/j.tate.2020.103075

Eddy-Spicer, D., Ehren, M., Bangpan, M., Khatwa, M., & Perrone, F. (2016). Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised? A realist synthesis of school accountability in low-and middle-income countries. *Social Science Research Unit, UCL Institute of Education, University College London*.

Ehren, M. (2024). *Conceptualising "accountability" for the comparative review.*

Ehren, M. (2025). *Context, mechanisms and outcomes of accountability systems in fourteen education systems. Working paper for the Centre for Education Systems*.

Ehren, M. C. M. (2025). Context, mechanisms and outcomes of accountability systems in fourteen education systems. In *Working paper for the Centre for Education Systems*.

Ehren, M. C. M., Altrichter, H., McNamara, G., & O'Hara, J. (2013). Impact of school inspections on improvement of schools-describing assumptions on causal mechanisms in six European countries. *Educational Assessment, Evaluation and Accountability*, *25*(1), 3–43. https://doi.org/10.1007/s11092-012-9156-4

Ehren, M. C. M., Gustafsson, J.-E., Altrichter, H., Skedsmo, G., Kemethofer, D., & Huber, S. G. (2015). Comparing effects and side effects of different school inspection systems across Europe. *Comparative Education*, *51*(3), 375–400.

Ehren, M. C. M., Janssens, F. J. G., Brown, M., McNamara, G., O'Hara, J., & Shevlin, P. (2017a). Evaluation and decentralised governance: Examples of inspections in polycentric education systems. *Journal of Educational Change*, *18*(3), 365–383.

Ehren, M. C. M., Janssens, F. J. G., Brown, M., McNamara, G., O'Hara, J., & Shevlin, P. (2017b). Evaluation and decentralised governance: Examples of inspections in polycentric education systems. *Journal of Educational Change*, *18*(3), 365–383. https://doi.org/10.1007/s10833-017-9297-9

Ehren, M., Hutchinson, J., & Muñoz-Chereau, B. (2023). Place-based disparities faced by stuck schools in England: a contextual understanding of low performance and the role of inspection outcomes. *School Effectiveness and School Improvement*, *34*(4), 401–418.

Ehren, M., Perryman, J., & Shackleton, N. (2015). Setting expectations for good education: How Dutch school inspections drive improvement. *School Effectiveness and School Improvement*, *26*(2), 296–327.

Eizadirad, A., Martinez, X., & Ruminot, C. (2016). Comparative analysis of educational systems of accountability and quality of education in Ontario, Canada and Chile: Standardized testing and its role in perpetuation of educational inequity. *Interfaces Brasil/Canadá*, *16*(2), 54–88.

Eren, O., Figlio, D., Mocan, N., & Ozturk, O. (2022). *School Accountability, Long-Run Criminal Activity, and Self-Su¢ ciency*.

European Commission. (2022). *Successful Pisa Stories in the EU*. https://doi.org/10.2766/429517

Evans, G. (2023a). Scratching beneath the surface: using policy archaeology to better understand education reform in Wales. *Management in Education*, 08920206231160646.

Evans, G. (2023b). *Understanding the impact of school inspection: An analysis of how school leaders in Wales interpret and respond to inspection outcomes*. University of Wales Trinity Saint David (United Kingdom).

Feng, L., Figlio, D., & Sass, T. (2018). School accountability and teacher mobility. *Journal of Urban Economics*, *103*, 1–17.

Figlio, D., & Loeb, S. (2011). School accountability. *Handbook of the Economics of Education*, *3*, 383–421.

Figlio, D. N., & Getzler, L. S. (2006). Accountability, ability and disability: Gaming the system? In *Improving school accountability* (Vol. 14, pp. 35–49). Emerald Group Publishing Limited.

Figlio, D. N., & Rouse, C. E. (2006). Do accountability and voucher threats improve low-performing schools? *Journal of Public Economics*, *90*(1–2), 239–255.

Figlio, D. N., & Winicki, J. (2005). Food for thought: the effects of school accountability plans on school nutrition. *Journal of Public Economics*, *89*(2–3), 381–394.

Friedman, M. (2002). *Capitalism and freedom Milton Friedman* (I. ebrary, R. D. Friedman, & R. D. Friedman, Eds.; 40th anniversary ed.). Chicago : University of Chicago Press, 2002.

Gardezi, S. (2024). The development and status of self-evaluation as a component of school inspection and quality assurance: A comparison across four countries. *Heliyon*, *10*(20). https://doi.org/10.1016/j.heliyon.2024.e39321

George, A. L., & Bennett, A. (2005). *Case studies and theory development in the social sciences*. Belfer Center for Science and International Affairs.

Gjefsen, H. M., & Gunnes, T. (2020). How school accountability affects teacher mobility and teacher sorting. *Education Economics*, *28*(5), 455–473.

Gregg, J. J., & Lavertu, S. (2023). Test-based accountability and educational equity: Breaking through local district politics? *Economics of Education Review*, *97*, 102485. https://doi.org/https://doi.org/10.1016/j.econedurev.2023.102485

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management*, *24*(2), 297–327.

Hanushek, E. A., Saenz-Armstrong, P., & Salazar, A. (2024). *Balancing Federalism: The Impact of Decentralizing School Accountability*. http://www.nber.org/papers/w32351

Hargreaves, A. (2020). Large-scale assessments and their effects: The case of mid-stakes tests in Ontario. *Journal of Educational Change*, *21*(3), 393–420.

Hashimoto, A. (2018). Note on the Characteristics of Japanese School Evaluation: Seven factors obstructing successful implementation. *Japanese Journal of Evaluation Studies*, *18*(2), 35–54.

Hayek, F. A. von. (2001). *The road to serfdom* (2nd ed.). Routledge.

Hoffer, T. B. (2000). Accountability in education. In *Handbook of the Sociology of Education* (pp. 529–543). Springer.

Holbein, J. B., & Ladd, H. F. (2017). Accountability pressure: Regression discontinuity estimates of how No Child Left Behind influenced student behavior. *Economics of Education Review*, *58*, 55–67.

Hong, M. K., Chen, F., Chen, Y.-Y., & Klenk, M. (2022). Supporting Qualitative Coding with Machine-in-the-Loop. *Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS 2022)*.

Honingh, M., van Genugten, M., de Gooyert, V., & Blom, R. (2022). What is the impact of a principle-based inspection approach? *School Effectiveness and School Improvement*, *33*(2), 236–259.

Hooge, E., Burns, T., & Wilkoszewski, H. (2012). *Looking beyond the numbers: Stakeholders and multiple school accountability*.

Huang, J., Yamada, H., & Zhou, Z. (2024). An analysis of the governance of the Japanese teacher license renewal system. *International Journal of Educational Research*, *128*, 102468.

Hughson, T., & Menzies, L. (2026). Polycentrism in Education Policy Making in England and Aotearoa New Zealand: Strengths as well as Weaknesses? In C. Skerritt (Ed.), *Opportunities and challenges for policy advice and policy making in education: Future possibilities*. Bloomsbury.

Hussain, I. (2015). Subjective performance evaluation in the public sector: Evidence from school inspections. *Journal of Human Resources*, *50*(1), 189–221.

Hwa, Y.-Y. (2021). Contrasting approaches, comparable efficacy?: How macro-level trust influences teacher accountability in Finland and Singapore. In *Trust, Accountability and Capacity in Education System Reform*. Taylor & Francis.

Hwa, Y.-Y. (2022). 'Our system fits us': comparing teacher accountability, motivation, and sociocultural context in Finland and Singapore. *Comparative Education*, *58*(4), 542–561.

Inspectorate of Education (Netherlands). (2025). *Inspection Framework 2021 for the supervision of secondary education*. https://english.onderwijsinspectie.nl/documents/publications/2025/09/04/inspection-framework-secondary-education-2021

Jacob, B. A., & Levitt, S. D. (2003). *Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating*.

Jang, E. E., & Sinclair, J. (2018). Ontario's educational assessment policy and practice: a double-edged sword? *Assessment in Education: Principles, Policy & Practice*, *25*(6), 655–677. https://doi.org/10.1080/0969594X.2017.1329705

Jisu, R. (2020). *Supporting Professional Capacity of Teachers Through Teacher Evaluation (Doctoral dissertation)*.

Jones, K. L., Tymms, P., Kemethofer, D., O'Hara, J., McNamara, G., Huber, S., Myrberg, E., Skedsmo, G., & Greger, D. (2017). 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*, *43*(6), 805–822.

Kane, T. J., & Staiger, D. O. (2002). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives*, *16*(4), 91–114.

Karsten, S. (1999). Neoliberal Education Reform in The Netherlands. In *Education* (Vol. 35, Issue 3).

Kay, A. (2012). Policy trajectories and legacies: path dependency revisited. In *Routledge handbook of public policy* (pp. 480–490). Routledge.

Keane, A. (2023). Changing features of school inspection: a perspective from Wales. *Forum*, *65*(3), 81–90.

Kim, K. (2024). *Teacher Decision-Making and Classroom Autonomy Under Test-Based Accountability: A Comparative Analysis*. The Florida State University.

Klimecki, R., & Wilson, D. (2025). Decentering governance models: education policy and school accountability in Post-Devolution Wales. *Critical Policy Studies*, *19*(3), 456–476.

Kogan, M. (1986). *Education accountability: An analytic overview*. Routledge.
Ladd, H. F. (2012). *School accountability: To what ends and with what effects?* https://cipe.umd.edu/conferences/EducationEvaluationItaly/COMPLETE_PAPERS/Ladd/LADD.pdf

Ladd, H. F., & Thompson, E. (2012). *School Accountability: To What Ends and With What Effects? Keynote address for Conference on "Improving Education through Accountability and Evaluation: Lessons from Around the World," sponsored by the Association for Public Policy Analysis and Management*.

Lee, Y.-J. (2024). The discursive construction of responsibility among teachers in Singapore:"Everything is good, so we do everything." *Asia Pacific Education Review*, *25*(1), 187–200.

Linn, R. L. (2008). Methodological issues in achieving school accountability. *Journal of Curriculum Studies*, *40*(6), 699–711.

Lodge, M. (2015). *Ian Ayres and John Braithwaite, responsive regulation: transcending the deregulation debate*.

Majone, G. (1997). From the positive to the regulatory state: Causes and consequences of changes in the mode of governance. *Journal of Public Policy*, *17*(2), 139–167. https://doi.org/10.1017/s0143814x00003524

Mansfield, J., & Slichter, D. (2026). The long-run effects of consequential school accountability. *Journal of Labor Economics*, *44*(1), 0.

Maroy, C., & Pons, X. (2021). Re-conceptualising education policy trajectories in a globalised world: lessons from a multi-level comparison of accountability in France and Quebec. *Comparative Education*, *57*(4), 560–578.

Maroy, C., & Voisin, A. (2017a). *UNESCO: think piece on accountability: background paper prepared for the 2017/8 Global education monitoring report.* https://doi.org/10.13140/RG.2.2.17900.18569

Maroy, C., & Voisin, A. (2017b). *UNESCO: think piece on accountability; Background paper prepared for the 2017/8 Global education monitoring report, Accountability in education: meeting our commitments; 2017*. https://doi.org/10.13140/RG.2.2.17900.18569

McArthur, D., & Reeves, A. (2022). The unintended consequences of quantifying quality: Does ranking school performance shape the geographical concentration of advantage? *American Journal of Sociology*, *128*(2), 515–551.

McClurg, K., Cantley, I., & Donnelly, C. (2025). The possibilities and problematics of school evaluation: post-primary teachers' perspectives on the impact of school inspections in a high stakes inspection system. *Irish Educational Studies*, *44*(3), 421–443.

McElroy, K. (2023). Does test-based accountability improve more than just test scores? *Economics of Education Review*, *94*, 102381.

McNamara, G., O'Hara, J., Brown, M., & Quinn, I. (2020). Quality assurance in Irish schools: Inspection and school self-evaluation. *Administration*, *68*(4), 161–180.

McNamara, G., Skerritt, C., O'Hara, J., O'Brien, S., & Brown, M. (2022). For improvement, accountability, or the economy? Reflecting on the purpose(s) of school self-evaluation in Ireland. *Journal of Educational Administration and History*, *54*(2), 158–173. https://doi.org/10.1080/00220620.2021.1985439

McNamara, Gerry, Brown, Martin, Gardezi, Sarah, O'Hara, Joe, O'Brien, Shivaun, & Skerritt, Craig. (2021). Embedding Self-Evaluation in School Routines. *Sage Open*, *11*(4), 21582440211052550. https://doi.org/10.1177/21582440211052552

Menzies, L. (2023). Continuity and Churn: Understanding and responding to the impact of teacher turnover. *London Review of Education*. https://doi.org/10.14324/LRE.21.1.20

Menzies, L., & Quilter-Pinner, H. (2023). *Improvement through Empowerment: Helping our teachers and schools be the best they can be*. https://ippr-org.files.svdcdn.com/production/Downloads/improvement-through-empowerment-nov23.pdf

Menzies, L., Yates, W., & Huband-Thompson, B. (2023). *Balancing Act: Navigating the Tensions in our School System*. https://ippr-org.files.svdcdn.com/production/Downloads/balancing-act-school-system-august-23.pdf

Millard, W., Small, I., & Menzies, L. (2017). *Testing the Water: How assessment can underpin, not undermine, great teaching*. The Centre for Education and Youth / Pearson. https://www.researchgate.net/publication/330812163_Testing_the_Water_How_assessment_can_underpin_not_undermine_great_teaching#fullTextFileContent

Møller, J., Hall, D., & Normand, R. (2025). Public education and teacher professionalism in an age of accountability. *Critical Studies in Education*, *66*(5), 664–682.

Moreton, H. J., Boylan, M., & Simkins, T. (2017). Headteachers who also inspect: Practitioner inspectors in England. In *School inspectors: Policy implementers, policy shapers in national policy contexts* (pp. 137–158). Springer.

Morozumi, A., & Tanaka, R. (2023). *School Accountability and Student Achievement: Neighboring Schools Matter*. Research Institute of Economy, Trade and Industry (RIETI).

Mulgan, R. (2000a). 'Accountability': An Ever-Expanding Concept? *Public Administration*, *78*(3), 555–573. https://doi.org/10.1111/1467-9299.00218

Mulgan, R. (2000b). 'Accountability': An Ever-Expanding Concept? *Public Administration*, *78*(3), 555–573. https://doi.org/10.1111/1467-9299.00218

Munoz Chereau, B., Hutchinson, J., & Ehren, M. (2022a). *"Stuck" schools: Can below good Ofsted inspections prevent sustainable improvement?*

Munoz Chereau, B., Hutchinson, J., & Ehren, M. (2022b). *'Stuck'schools: Can below good Ofsted inspections prevent sustainable improvement?*

Nunes, L. C., Reis, A. B., & Seabra, C. (2015). The publication of school rankings: A step toward increased accountability? *Economics of Education Review*, *49*, 15–23.

OECD. (2018). *TALIS 2018 Results (Volume II) Annex C Table II.3.30*.

Ofsted. (2023). *Education inspection framework for September 2023*. GOV.UK. https://www.gov.uk/government/publications/education-inspection-framework/education-inspection-framework-for-september-2023

O'Mara-Eves, A., Vigurs, C., Ko, S. Y. (Jessica), & Simon, A. (2025). *Rapid systematic evidence review of school accountability policies in secondary education across fourteen jurisdictions*. EPPI Centre.

O'Mara-Eves, A., Simon, A., Anand, K., Stansfield, C., Vigurs, C., & Ko, S. Y. (Jessica). (2025b). *Technical Report for Accountability Policy in Secondary Education. EPPI Centre*.

Ontario Ministry of Education. (2013). *Policy/Program Memorandum No. 119: Developing and implementing equity and inclusive education policies in Ontario schools*. Government of Ontario. https://www.ontario.ca/document/education-ontario-policy-and-program-direction/policyprogram-memorandum-119

Pawson, R. (2006). *Evidence-based policy: a realist perspective*.

Pawson, R. (2024). *How to think like a realist: A methodology for social science*. Edward Elgar Publishing.

Pinto, L. E. (2016). Tensions and fissures: The politics of standardised testing and accountability in Ontario, 1995–2015. *The Curriculum Journal*, *27*(1), 95–112.

Poisson, M., & Jorgoni, E. (2025). *OPEN SCHOOL DATA IN EUROPEAN EDUCATION SYSTEMS Council of Europe Platform on Ethics, Transparency and Integrity in Education (ETINED)-Volume 9 Council of Europe*. http://book.coe.int

Priestley, M. (2013). Curriculum regulation in Scotland: A wolf in sheep's clothing is still a wolf. *European Conference on Curriculum Studies*.

Proudfoot, K. (2025). Creeping anglicisation: school inspection in Scotland and Ireland. *Comparative Education*, 1–18.

Reback, R., Rockoff, J., & Schwartz, H. L. (2014). Under pressure: Job security, resource allocation, and productivity in schools under No Child Left Behind. *American Economic Journal: Economic Policy*, *6*(3), 207–241.

Rezai-Rashti, G. M., & Segeren, A. (2023). The game of accountability: perspectives of urban school leaders on standardized testing in Ontario and British Columbia, Canada. *International Journal of Leadership in Education*, *26*(2), 260–277.

Rockoff, J., & Turner, L. J. (2010). Short-run impacts of accountability on school quality. *American Economic Journal: Economic Policy*, *2*(4), 119–147.

Rouse, C. E., Hannaway, J., Goldhaber, D., & Figlio, D. (2013). Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure. *American Economic Journal: Economic Policy*, *5*(2), 251–281.

Sahlberg, P. (2016). The global educational reform movement and its impact on schooling. *The Handbook of Global Education Policy*, 128–144.

Seawright, J., & Gerring, J. (2008). Case selection techniques in case study research: A menu of qualitative and quantitative options. *Political Research Quarterly*, *61*(2), 294–308. https://doi.org/10.1177/1065912907313077

Şen, A. Y., & Maya, İ. (2022a). School quality assurance applications for basic education schools. *Education Quarterly Reviews*, *5*.

Şen, A. Y., & Maya, İ. (2022b). School quality assurance applications for basic education schools. *Education Quarterly Reviews*, *5*.

Senedd Research. (2025). *What's happening to school improvement services?* https://research.senedd.wales/research-articles/what-s-happening-to-school-improvement-services

SICI. (2024). *Country Profile Dutch Inspectorate of Education*. www.sici-inspectorates.eu/swfiles/files/SICI_-_Country_Profile_Template_the-Netherlands_27022024-1.pdf

Sinclair, A. (1995). The chameleon of accountability: Forms and discourses. *Accounting, Organizations and Society*, *20*(2–3), 219–237.

Smith, A. [1776]. (2005). *An inquiry into the nature and causes of the wealth of nations: a selected edition*. AN ELECTRONIC CLASSICS SERIES PUBLICATION. https://www.rrojasdatabank.info/Wealth-Nations.pdf

Smith, W. (2025). Understanding accountability in education: a review of common typologies. In *Handbook on Comparative Education* (pp. 252–264). Edward Elgar.

Tian, M. (2025). Embracing complexity: rethinking education inspection in England. *Educational Review*, *77*(5), 1500–1516. https://doi.org/10.1080/00131911.2024.2325660

Torres, R. (2021). *Does test-based school accountability have an impact on student achievement and equity in education?: A panel approach using PISA*.

Udahemuka, M. (2016). *Signal loss: What we know about school performance*. New Zealand Initiative.

Wang, F. (2017). The Ontario quandary: Principals' perceptions of accountability policy and social justice. *Journal of School Leadership*, *27*(3), 392–420.

West, A., Mattei, P., & Roberts, J. (2011). Accountability and Sanctions in English Schools. *British Journal of Educational Studies*, *59*(1), 41–62. https://doi.org/10.1080/00071005.2010.529416

Westhorp, G., Walker, B., Rogers, P., Overbeeke, N., Ball, D., & Brice, G. (2014). Enhancing community accountability, empowerment and education outcomes in low and middle-income countries: A realist review. *EPPI-Centre, Social Science Research Unit, Institute of Education, University of London*.

White Paper: Educational Disadvantage and the Educational Needs of Immigrants (1974).

Wittgenstein, L. (2009). *Philosophical investigations*. John Wiley & Sons.

Wöbmann, L., Lüdemann, E., Schütz, G., & West, M. R. (2007). *School accountability, autonomy, choice, and the level of student achievement: International evidence from PISA 2003*.

World Bank. (2008). *What is school-based management?* World Bank. http://documents.worldbank.org/curated/en/113901468140944134