Educational Research and Innovation

Governing Education in a Complex World

Edited by Tracey Burns and Florian Köster

Centre for Educational Research and Innovation

OECD

Educational Research and Innovation

Governing Education in a Complex World

Edited by
Tracey Burns and Florian Köster

OECD
BETTER POLICIES FOR BETTER LIVES

Please cite this publication as:
Burns, T. and F. Köster (eds.) (2016), *Governing Education in a Complex World*, Educational Research and Innovation, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264255364-en

ISBN 978-92-64-25535-7 (print)
ISBN 978-92-64-25536-4 (PDF)

Series: Educational Research and Innovation
ISSN 2076-9660 (print)
ISSN 2076-9679 (online)

Photo credits: Cover © Juriah/123RF.COM.

Corrigenda to OECD publications may be found on line at: *www.oecd.org/about/publishing/corrigenda.htm*.

Foreword

One of the crucial issues for OECD countries is how to deliver high quality, efficient, equitable and innovative education in increasingly complex education systems. A number of intersecting trends contribute to this increasing complexity: decentralisation has allowed local authorities, school boards and schools a greater degree of freedom to respond to local demands. Parents in OECD countries have become more diverse, individualistic and highly educated. With more readily available evidence about school and student achievement, stakeholders have also become more demanding towards schools to cater to students' individual needs. Education systems are now characterised by multi-level governance where the links between multiple actors operating at different levels are to a certain extent fluid and open to negotiation.

The Centre for Educational Research and Innovation (CERI)'s Governing Complex Education Systems (GCES) project focuses on which models of governance are effective in complex education systems and which knowledge systems are needed to support them. Its focus on complexity is connected to a broader organisational reflection on New Approaches to Economic Challenges (NAEC), which seeks to renew and strengthen the OECD's analytical frameworks, policy instruments and tools. A key element of this reflection is understanding the complex and interconnected nature of the global economy to allow for identifying synergies (e.g. between growth, inequality, stability and the environment) and strengthening the ability to manage policy trade-offs.

The GCES project has identified three themes vital for effective governance and successful reform: accountability, capacity building and strategic thinking. Accountability addresses the challenge of holding different actors at multiple levels responsible for their actions. Capacity building focuses on identifying gaps, skill needs and dynamics of implementation on individual, institutional and system level. Strategic vision pertains to the development of a long-term plan and set of common goals for the educational system among a broad array of actors. It requires aligning the different perspectives and time-horizons so that everyone involved can act together.

Creating the open, dynamic and strategic governance systems necessary for governing complex systems is not easy. *Governing Education in a Complex World* challenges our traditional concepts of education governance through work on complexity, change and new modes of decision-making. In doing so it sets the agenda for thinking about inclusive, adaptable, and flexible accountability and governance, necessary for governing complex systems in today's world. It offers examples from Austria, England (United Kingdom), the Netherlands and the United States, and ends with a suggestion for a way forward.

This publication is the first volume in a set of two. The second volume synthesising the findings from the six case studies carried out in the Governing Complex Education Systems project will be published later in 2016.

This publication was edited by Tracey Burns and Florian Köster of the Centre for Educational Research and Innovation (CERI) at the OECD. Within the OECD Secretariat Célia Braga-Schich, Sophie Limoges, Leonora Lynch-Stein and Anne-Lise Prigent provided valuable editorial support.

Acknowledgements

The Governing Complex Education Systems project, spanning from 2011 to 2016, would not have been possible without the support of a large number of individuals and countries. First, we wish to acknowledge the hosts of the thematic conferences that helped define and develop our analytical agenda: Oslo (Norwegian Ministry of Education and Research), The Hague (Dutch Ministry of Education and Research), Warsaw (Polish Ministry of Education), Paris (OECD and UNESCO) and Tallinn (Estonian Ministry of Education and Research). We thank the participants from 30 different countries who took part in the meetings and shared their governance challenges and successes with us and their colleagues.

We would also like to give a special mention to Helen Ängmo, Gábor Halasz, Jan Herczynski, Ülle Kikas, Peter Klandt, Ted Reininga, Rien Rouw, Micheline Scheys, Petter Skarheim, and Eli Sundby for their continued support and encouragement of this work from its inception.

The editors would like to thank the authors who contributed to this volume (in order of their chapters) Mark Mason, William C. Smith, Henno Theisens, Edith H. Hooge, Lorenz Lassnigg, Philippa Cordingley, Patrick Blanchenay and Lucie Cerna.

We also express our gratitude to team members (past and present) Patrick Blanchenay, Marlon Brandt, Lucie Cerna, Alina Kleinn, Maaria Klemola, Sonia Kosunen, Élodie de Oliveira, Mikko Silliman, William Smith and Sean Snyder for their work in this project over the years. An enormous thank you to Henno Theisens, who originally designed and proposed the Governing Complex Education Systems project, and Harald Wilkoszewski, a founding team member. We could not have done it without you.

In addition, we would like to extend our thanks to the many colleagues within the OECD Secretariat who also supported and developed this work through peer reviews of working papers and project proposals, particularly Stuart Elliot, Sonia Guerriero and Dirk Van Damme.

Tracey Burns and Florian Köster, Editors

Table of contents

Figures

Tables

Executive Summary

Governing multi-level education systems effectively requires governance models that balance responsiveness to local diversity with the ability to ensure national objectives. This is a delicate equilibrium, one that is difficult to achieve given the complexity of the education system in many OECD countries. As a result, governance issues have moved up political and policy agendas, and countries are increasingly looking for models that they can adapt to their own needs.

A number of intersecting trends contribute to increasing complexity in education systems. Decentralisation has allowed local authorities, school boards and schools a greater degree of freedom to respond to diverse and local demands. Parents, students and communities have become more diverse and highly educated and as a result have higher expectations that schools cater to students' individual needs. More information about school and student achievement is readily available, empowering a broader set of actors. This has changed the nature of the relationship between governance levels, moving away from a hierarchical relationship to a division of labour, interdependence and self-regulation. Education systems are now characterised by multi-level governance where the links between multiple actors operating at different levels are to a certain extent fluid and open to negotiation.

Governing Education in a Complex World addresses key challenges involved in governing modern education systems, looking specifically at complexity, accountability, capacity building and strategic thinking. The publication brings together research from the OECD Secretariat and invited chapters from international scholars to provide a state of the art analysis and a fresh perspective on some of the most challenging issues facing educational systems today.

Setting the stage: Governance in complex systems

Effective modern governance requires coordinated system-wide change involving a broad set of actors

Part 1 explores the concept of complexity and discusses its implications for educational governance. Chapter 1 provides an overview of the volume and proposes a set of key principles for modern governance of complex education systems. Chapter 2 looks at complexity theory and the argument that a significant degree of complexity in a system – whether an education system or a school – leads to emergent properties beyond those predictable from initial conditions. The discussion explores preconditions required to generate sustainable, positive, system-wide change in education. Chapter 3 explores the potential of governance networks, in which coordinated changes replace isolated interventions and align reforms to a system's contexts and resources.

Accountability and stakeholder involvement in complex systems

A constructive accountability system that balances the monitoring and pressure required to ensure efficient system functioning with support for improvement is crucial

Part 2 looks at the fundamental role of accountability in governance. Chapter 4 examines the increased use of test scores for accountability purposes worldwide, and asks how national testing systems are related to improvements in education performance. Chapter 5 discusses trends in accountability mechanisms and argues that regulatory and school performance accountability (vertical accountability), can be usefully augmented by horizontal measures involving multiple stakeholders. This has the potential to ultimately improve the level of achievement as well as the quality of education.

Capacity building and the use of knowledge

Local capacity building must be aligned with system vision and include a greater role of practitioners in knowledge production

Part 3 focuses on capacity building and the use of knowledge in increasingly complex education systems, providing concrete country examples. Chapter 6 examines the example of Austria and gives an outlook on how local capacity building could help mitigate governance inefficiencies in structurally complex systems. Chapter 7 looks at the role of knowledge production and knowledge use among teachers in improving instruction, providing concrete examples from England. The chapter discusses how policy can build capacity for teachers' involvement in research by providing tools and connecting points with large-scale education research.

Innovative and strategic governance: the role of policy experimentation and risk-taking

Taking risks and learning from failure is essential for innovation and the evolution of education systems. Controlled experimentation is one way forward

Part 4 explores tools for approaching reform. Policy experimentation has the potential to be an effective instrument for policy making in a complex environment but faces a number of difficulties in moving from theory to practice. Chapter 8 explores ecosystem experimentation as a way to operate at a small enough level to safeguard quality while maintaining a wide enough scope to allow for translating the findings to the system level (scaling-up). Chapter 9 discusses the Netherlands' experience in policy experimentation and the insights gleaned over the past decade. Chapter 10 turns to risk-taking, an often over-looked but essential part of any governance system as it is vital for innovation. Modern education systems have to allow for risk-taking, acknowledging the inherent possibility of failure and build processes to learn from failures.

The way forward: lessons learnt, and one essential ingredient: trust

Legitimate trust is a cornerstone of effective and efficient governance

The last chapter concludes the volume with the lessons learned from the work of the OECD on complex education governance. It also links the discussion to an often overlooked point: successful modern governance is built on trust. The final chapter examines this seemingly simple, yet decidedly complex topic in relation to the main themes in this volume. It ends with a return to the key principles of modern educational governance.

Creating the open, dynamic and strategic governance systems necessary for governing complex systems is not easy. This volume challenges our traditional concepts of education governance through work on complexity, change/reform and new modes of collaborative networks and decision-making. In doing so it sets the agenda for thinking about inclusive, adaptable, and flexible accountability and governance, necessary for governing complex systems in today's global world.

PART 1.

SETTING THE STAGE: GOVERNANCE IN COMPLEX SYSTEMS

Chapter 1.

Modern governance challenges in education

Tracey Burns and Florian Köster

Centre for Educational Research and Innovation, OECD

Education systems now tend to be characterised by multi-level governance where the links between multiple actors operating at different levels are more fluid and open to negotiation. As a result, the governance of complex multilevel education systems has become a policy priority. This chapter sets the stage for the publication by exploring the concept of complexity and its implications for modern education governance. It then provides the reader with an overview of the key themes of governing complex education systems – accountability, capacity building and strategic thinking. It sets out a set of principles for strategic thinking and modern governance, developed through OECD work with countries. The chapter concludes with an overview of the full volume, as well as a reminder of one currently under-studied issue that is the glue of modern governance: trust.

Introduction

Complexity in education systems is on the rise due to a number of intersecting trends. Parents in OECD countries have become more diverse, individualistic and highly educated. As evidence about school and student achievement has become more readily available, parents and other stakeholders have also become more demanding, pushing schools to cater for students' individual needs.

One of the most important responses to this increasing complexity has been decentralisation: allowing local authorities, school boards and schools a greater degree of freedom to respond to diverse and local demands. In fact, decentralisation may be too limited a term for what has happened. In many countries tasks have not simply devolved to regional, local or school levels. Lump sum funding, strengthening of stakeholders, horizontal accountability and holding local authorities and schools accountable through performance indicators have changed the nature of the relationship between the central, regional and local levels, moving away from a hierarchical relationship to a division of labour and more mutual independence and self-regulation. Education systems are now characterised by multi-level governance where the links between multiple actors operating at different levels are more fluid and open to negotiation.

These developments have been taking place in all OECD countries to varying degrees in the past three decades. Of course different countries started at different points of departure. Federal states, such as Australia, Austria, Canada, Germany or the United States, have the added complexity of authority spread over national and state levels. Certain countries have a long tradition of strong decentralisation (e.g. Finland, the United Kingdom). Other countries have a lengthy practice of freedom of school choice and of the establishment of publicly funded private schools (like Belgium or the Netherlands). These different points of departure are important for structural as well as traditional reasons, and have a great impact on the types of policy options available in that context.

Whatever the precise structure of their education systems, many OECD countries have been searching for governance models that allow them to effectively steer complex education systems. This search has led to a multiplication of governance mechanisms that are often applied simultaneously. For example, central ministries act as regulator for the education system, setting the rules within which increasingly autonomous schools must operate. But ministries also act as top-down enforcers of quality standards if schools consistently fail to meet expectations. Crucially, ministries are no longer the only actor involved in governing education systems. Apart from the increased role for schools and local administrations, there is a host of other stakeholders (including teacher unions, other ministries and national boards, teachers, parents, the media and students themselves) that play a role. When it comes to national strategy setting, negotiation and dialogue have become important governance mechanisms.

While decentralisation and the introduction of new governance mechanisms is an understandable and common response to complexity, they also further contribute to the complexity of the system. And despite all these changes, one element has persisted: ministries of education remain responsible for ensuring high quality, efficient, equitable and innovative education at the national level. They must fulfil this responsibility at the same time as the system engages with more diverse local actors, strong parental voice, higher levels of school autonomy, and newly important players like the media.

This chapter sets the stage for the publication by exploring the concept of complexity and its implications for modern education governance. It then provides the reader with an overview of the key themes of governing complex education systems – accountability, capacity building and strategic thinking. The chapter concludes with an overview of the full volume and the individual chapters of each of the contributing authors.

Two key questions

Governing multi-level education systems effectively requires governance models that balance responsiveness to local diversity with the ability to ensure national objectives. This is a delicate equilibrium and very difficult to achieve given the complexity of the education system in many OECD countries. As a result, governance issues have moved up on the agenda, and countries are increasingly looking for examples of good practice and models that they can adapt to their own needs. This policy priority led to the creation of an OECD/CERI project, *Governing Complex Education Systems*, or GCES.[1] The present volume emerges from the work of this project.

Box 1.1. The Governing Complex Education Systems (GCES) project

Launched in 2011, the OECD/CERI Governing Complex Education Systems project had the following three goals:

- Establish the state of research and evidence in governance of education systems and use of knowledge and contribute to the analytical and conceptual knowledge base in the field.
- Explore current practices in OECD member countries through a series of thematic workshops, working papers and case studies.
- Build an international network of policy makers and researchers with expertise in this area.

To this end, the project organised a series of thematic conferences to build an international network and bring together relevant stakeholders from policy, research, and practice. It produced a range of working papers exploring the conceptual issues around modern governance challenges. A series of case studies from Belgium (Flanders), Germany, the Netherlands, Norway, Poland and Sweden provided an empirical investigation of key issues of multi-level education governance. The project's work culminates in two stand-alone volumes: the present publication *Governing Education in a Complex World* and *Governance in Action: Synthesis of Case Studies*, which compares and integrates the findings of the six case studies carried out over the course of the GCES project.

Annex 1.A1 presents the full list of project outputs including conferences, working papers and case studies.

Addressing the search for adaptable examples of effective governance systems, the first key question of the project, and thus our discussion, was:

1. *"What models of governance are effective in complex education systems?"*

For the policy maker tasked with developing a response to a particular issue, it is often not fully clear what kinds of evidence are needed in order to address key policy issues – and in fact there may be multiple paths to a particular evidence-based solution. Policy makers must build a repertoire of actions and strategies to navigate the knowledge options available. Apart from the use of knowledge by policy makers, important

questions are how and where knowledge is produced and how it is transmitted to policy makers. In this context, the role of brokerage agencies in providing timely evidence and helping weigh the various options available are becoming increasingly important. This raises the second key question for modern governance:

2. *"What knowledge system is necessary to support the effective governance of complex education systems?"*

An important component of modern governance systems is their capacity to learn and to share knowledge. With the growth in complexity, governance has become a knowledge intensive activity. In complex and often fragmented systems, sharing knowledge between different parts of the system is essential, for example, to make innovative practice at decentralised levels available in other (decentralised) parts of the system. The key role of knowledge becomes more important as the different types of testing and assessment on national and international levels have led to an explosion in the kinds and types of evidence available to policy makers. Of course, knowledge is also generated by professional experience and includes tacit knowledge transmitted informally within systems.

Knowledge and learning are also vital elements in negotiations and dialogue that are essential to creating consensus in complex systems with multiple actors. Knowledge becomes a tool to steer the system: providing decentralised decision makers and practitioners with relevant, high quality knowledge is imperative to improve the quality of decision making and practices.

Box 1.2. Why governance and knowledge, and why now?

The OECD Secretary-General has recently proposed a reflection process to explore New Approaches to Economic Challenges (NAEC), which aims to revisit and assess whether analytical frameworks and economic models need to be adapted to a post-crisis world. A key issue for this reflection is the concept of a Strategic State:

It is not so much the size of the State which is at stake, but rather its governance. In other words, it is not so much a reduced state that we need to foster economic growth in our countries, but a strategic state. This idea of a strategic state that targets its investments to maximize growth in the face of hard budget constraints departs both, from the Keynesian view of a state sustaining growth through demand-driven policies, and from the neoliberal view of a minimal state confined to its regalian functions (public order, basic services). (Aghion, 2012)

One of the key themes of this work is the impact of the crisis on trust in government. Dramatic cuts in social expenditure have raised concerns about fairness, equity of sacrifice, and worries about the social contract. As governments struggle to communicate a clear vision for recovery, the public's trust in government must be reinforced, and efforts must be made to strengthen institutions and build capacity across different dimensions of trust (e.g. reliability, fairness and impartiality, integrity and honesty, and inclusiveness) (OECD, 2015a).

At the same time, the concept of a *smarter state* includes a focus on government learning. Although traditionally thought of in terms of innovation and industrial policy, this concept extends to all sectors of government and includes an emphasis on trying new approaches, learning what works, and building the systemic capacity of the government to improve policy design, steering, and implementation. Finding new approaches to economic challenges, then, requires revisiting governance models in all areas, including education.

Modern governance and the complexity challenge

Navigating modern governance requirements is easier said than done. There are no magic solutions, no one-size-fits-all recipe that can be rolled out to guarantee success. The multitude of possible solutions to any given problem can be bewildering; and it is certainly frustrating to any politician looking for fast answers. It has been argued that the one constant in education governance is surprise: "At any given moment, there is a high probability of low probability events. In other words, surprise dominates" (Dror, 1986). Why would this be so? One hint is that education systems are complex systems, and thus are not easily governed by linear logic and processes (Snyder, 2013).

Defining complex systems

Our world is becoming more complex, with more dynamic growth and interaction in worldwide trends than ever before. Complexity theory posits that systems begin as collections of individual actors who organise themselves and create relationships. These relationships form in response to positive or negative feedback, as well as a degree of randomness. New structures and behaviours then emerge as the actors act and react to each other. A complex system has the following core components (Sabelli, 2006):

- Behaviour is not explained by the properties of the components themselves, but rather emerges from the interaction of the components.
- The system is non-linear and relies on feedback to shape its evolution.
- The system operates on multiple time-scales and levels simultaneously.

Analytically, complex systems pose several challenges as a particular system can no longer examined in isolation. Rather, the study of complex systems requires a step back to look at how the various interconnections can form a coherent whole.

What makes complex problems unique?

In order to address governance issues from the perspective of complexity, it is useful to distinguish between simple, complicated and complex problems (Glouberman and Zimmerman, 2002; see also Table 1.1).

Table 1.1. Simple complicated and complex problems

Simple: Following a recipe	Complicated: Sending a rocket to the moon	Complex: Raising a child
Recipes are essential	Formulae are critical	Formulae have limited application.
Recipes are easily replicated	Sending one rocket increases assurance that the next will be ok	Raising one child gives experience, but no assurance of success with another
Expertise is helpful but not required	High levels of expertise in multiple fields needed	Expertise can contribute but is neither necessary nor sufficient for success
Produces a standardised product	Rockets are similar in critical ways	Each child is unique and must be approached individually
Best recipes give good results every time	There is a high degree of certainty in the outcome once the original issues are solved	Uncertainty of outcome remains

Source: Snyder (2013); adapted from Glouberman and Zimmerman (2002).

Educational governance often attempts to follow a complicated approach when developing solutions to complex problems (Duit et al., 2010, see also Mason [this volume]). As Johnson (2008) argues:

> Currently, many methods of investigating the educational outcomes of individual schools […] are based on linear algorithms that simplify and break down systems into isolated component parts. The premise of such linear models is that inputs into the system will result in predictable outcomes. While appropriately predictive of some static, closed systems, these models fail to adequately predict the behaviour of or capture the essence and emergent properties of complex systems involving three or more interacting components. (Johnson, 2008: 5-6, cited in Snyder [2013])

Dimensions of complexity play a major role in how, and in what ways, education might be effectively governed. Modern education governance must be able to juggle the dynamism and complexity at the same time as it steers a clear course towards established goals. And it must do this as efficiently as possible, with limited financial resources. Education systems are complex in at least the following dimensions:

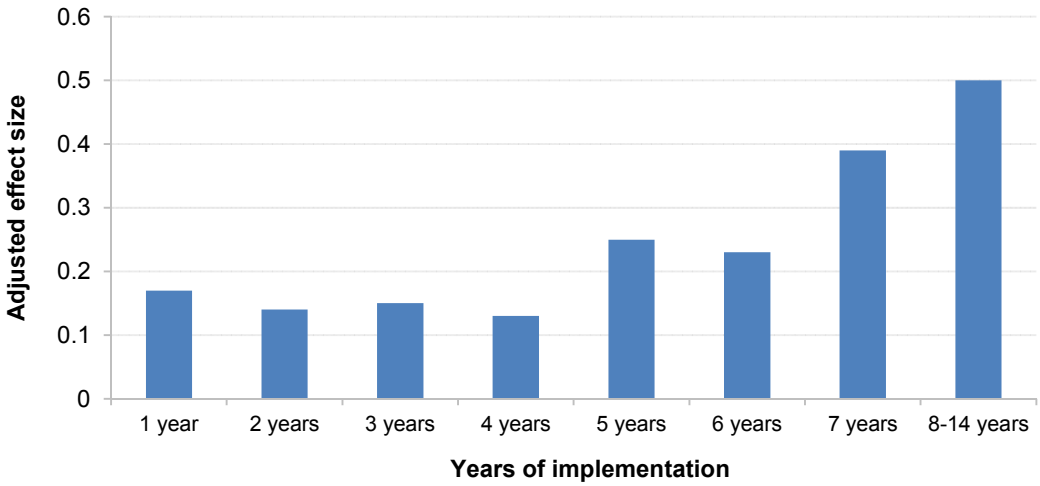
- They are multi-level systems (local, regional, national in many countries) and alignment is a major challenge, particularly in those most decentralised (Hopfenbeck et al., 2013; Blanchenay, Burns and Köster, 2014).
- Reflecting our societies, they are increasingly diverse – both in terms of the demographics of the population (of students, of teachers, and communities) as well as the values and identities we ascribe to ourselves and expect our education systems to deliver.
- They contain a growing number of stakeholders who are increasingly vocal about their wants and desires, not only for themselves and their children, but for the systems as a whole.
- Education is a field with strong a priori beliefs, strongly tied both to our identities and our experiences. Not only do we expect education to deliver the kinds of citizens we desire, everyone has taken part in education in some form or another. In doing so they have often formed strong personal opinions about what appears to work, and what does not, and these opinions may not be aligned with research findings.

The reality of modern governance

This complexity in the system is matched by new governance challenges in our modern world. Governance and political life is more and more marked by turbulence and surprises, and there is a growing cynicism about government and public institutions in general. Part of this is due to decreasing levels of trust, especially of our elected leaders (Cerna, 2014), and lasting impacts from the financial crisis of 2007-08. But part of this is also due to new expectations of governance, where an emphasis is placed on simple, fast and effective (although possibly not lasting) solutions.

In this world marked by new technologies and instant feedback, expectations tend to rise faster than performance. This is not particularly surprising, given the time it takes to see the effect of an educational reform: one meta-analysis of broad compulsory school reforms in the United States suggests that the strongest effects are seen 8 to 14 years after a reform is begun (Borman et al., 2003, see Figure 1.1).

Figure 1.1. Effect size of compulsory school reform, by years of implementation



Note: Effect sizes based on meta-analysis of studies pertaining to the impact of comprehensive school reform on student achievement in the United States.

Source: Borman et al. (2003).

This can become a challenge in terms of responding to the needs of increasingly vocal stakeholders. Parents concerned about the education of their children do not have eight years to wait and see what is effective; they expect the best education for their children now. The timeline is also completely at odds with the needs of an elected official, who has two or three years to make his/her case for re-election and demonstrate the efficacy of any flagship reforms.

Time, then, is another element of complexity. Elections at multiple levels create short-termism at the same time that research has demonstrated that the effects of a reform can take a significant amount of time to bear fruit. Further adding to the complexity, the effects of a reform may in fact change over time: in the realm of school choice for example, Waslander et al. (2010) point out that short term effects (generated by the early, generally well-informed adopters of a policy) can be quite different than longer term effects (when more parents have had a chance to act on it).

So what can be done? The answer to this lies in the answer to the second question posed earlier in the Introduction: *What knowledge system is necessary to support the effective governance of complex education systems?* Work on this element has made clear that the necessary knowledge system needs to build on rich and nuanced data that are also easily understandable. In fact, the necessary knowledge system combines descriptive system data (on achievement, graduation, etc.) with research findings that can determine whether something is working, and why. It also includes the wealth of practitioner knowledge available, both formalised and informal. The key is to knowing what to use, when, and why (Fazekas and Burns, 2012). And in a complex system marked primarily by surprises, this is no easy task.

As described above, modern governance must take into account the complexity of our systems, as well as the major themes that countries struggle with: accountability, capacity building, and strategic thinking. All of these issues must also be tied into the human element, and trust, which is the backbone of any functioning governance system. The role

of knowledge and evidence in guiding strategic decisions underlies all of these elements. The following section explores each of these themes in turn.

Accountability, capacity building and strategic thinking

Accountability

The issue of accountability is central to the governance of complex education systems, especially in terms of setting priorities and steering in multilevel systems with sometimes overlapping actors. Accountability gaps, for example situations where the central level may no longer be the driving force for accountability purposes but there is not a clear or functioning replacement, are one of the challenges that many countries face. There is a very real question about which actors at which levels should be held accountable for which outcomes, and how this can function in a coherent and intelligent manner.

In this context the role of evaluation and, specifically, of performance measurement, in managing accountability in the system is an important factor. In education, as in many other public sectors, a stronger focus on measureable outputs (for example, student test scores, graduation rates and transitions to the labour market) has been accompanied by an increased emphasis on standardised comparable testing. This has led to an explosion of evidence available to policy makers (and indeed all stakeholders in those countries where achievement data is made publicly available to all stakeholders). This has been an important force in increasing the transparency and accountability of the system, and in helping to identify areas for improvement.

However, there have been some perhaps unintended drawbacks to increasing accountability. There is an inherent tension between accountability and innovation, in that tightly controlled accountability mechanisms seek to minimise risk and error, both of which are fundamental elements of the innovation process. Yet countries are under strong internal (and at times external) pressure to strengthen their accountability systems while at the same time encouraging innovation. These kinds of inherent tensions are part and parcel of modern governance challenges, as they require a) a systemic vision that can identify the tension and b) making an informed choice about the best way to balance the competing elements in a particular system or school.

It is clear that all countries would like to have a strong accountability system that increases achievement and excellence while at the same time allowing for creativity and innovation. It is thus necessary to move away from thinking about effective accountability as simply implying more evaluation and mechanisms to ensure compliance. Rather, OECD countries are now at the stage where they are thinking systematically about their goals and desires for accountability as a tool for improvement, which includes also the room to innovate on all levels including the school and classroom (OECD, 2013). We argue that the term *strong accountability* should thus entail an explicit acknowledgement of the complexity involved and the need for a constructive approach that includes an understanding of the balance between regulations and evaluation instruments on the one hand and other elements of education excellence and equality, such as space and time to study subjects that are not part of national tests, or the participation and feedback of a wide variety of educational actors, on the other.

Strong accountability systems thus keep a clear focus on achievement and excellence, while being nuanced enough to allow for innovation, creativity and a rounded learning experience. This requires balancing evaluation and assessment with the risk-taking and

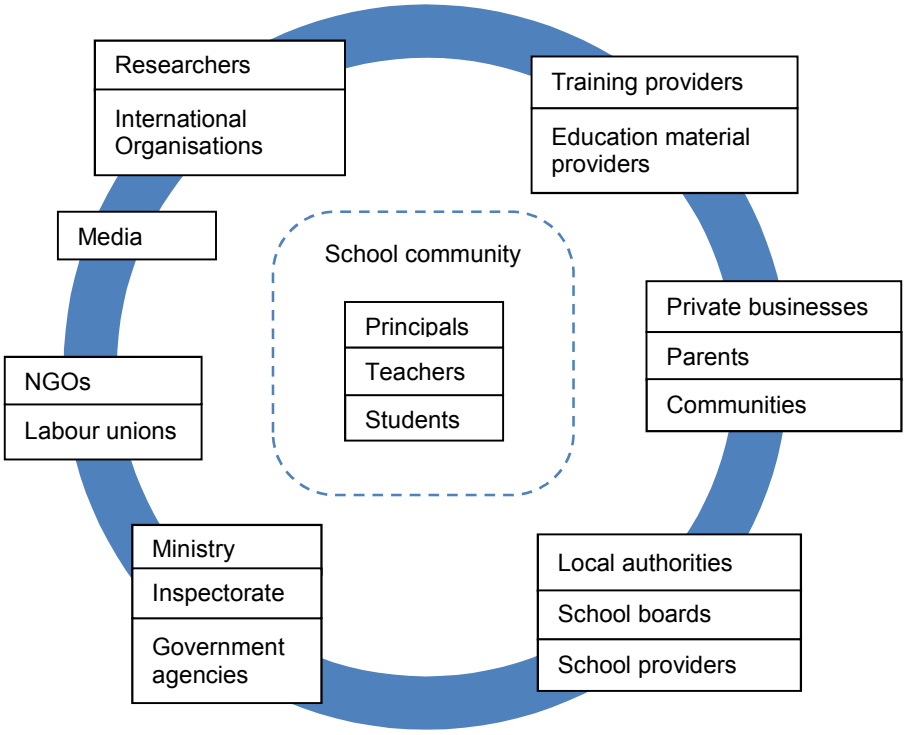
potential failure involved in innovation, both on the level of the practitioner as well as the system. The exact configuration of an education system's accountability system will of course depend on context and how decision-makers (and communities) choose to balance these various processes. It will also depend on the level and extent of stakeholder engagement and activity in the governance process.

Modes of accountability

Two types of accountability mechanisms are commonly used: vertical and horizontal. *Vertical accountability* is generally top-down and hierarchical. It enforces compliance with laws and regulation and/or holds schools accountable for the quality of education they provide. *Horizontal accountability* presupposes non-hierarchical relationships. It is directed at how schools and teachers conduct their practice and/or at how schools and teachers provide multiple stakeholders with insight into their educational processes, decision-making, implementation and results (see also Hooge, Chapter 5 of this volume, for a more detailed discussion).

Hooge et al. (2012) argue that vertical measures of accountability, that is, more traditional regulatory and school performance accountability, can be usefully augmented by horizontal measures involving multiple stakeholders. These would include actors such as parents, students, and communities (see Figure 1.2). Systems of multiple school accountability aim to efficiently and effectively take into account the nuanced nature and purposes of education, including an openness to innovation and creativity in multiple subject areas.

Figure 1.2. Potential stakeholders in education



But do these accountability mechanisms really work? Reports are mixed: there is great appreciation for the process and a broader range of stakeholder voices. However, ministries report a reluctance to rely too heavily on information generated by multiple accountability mechanisms due to doubts about its reliability and the risk of information overload. On this basis, central government is advised to discuss the purposes and use of multiple accountability mechanisms with the institutions and to balance the opportunities (information to learn, improve, steer, and formulate policies) with the risks (e.g. information overload) (Hooge et al., 2012).

In education, multiple accountability is still a fairly new concept and the amount of available research on how to make it work is modest. Three lessons, however, can be learned from existing models in the Netherlands and the United Kingdom (as described in Hooge et al., 2012):

- *Identify the key stakeholders.* This is more difficult than it sounds, and schools must make efforts to involve less powerful or inactive voices.
- *Build capacity for this new role.* Some stakeholders might not have the knowledge and language needed and may inadvertently be excluded in accountability processes. Providing them with the tools to interpret and analyse benchmarking data and other evaluation processes (e.g. value-added measures) is an important part of giving them the expertise they need to take part.
- *Schools need to be ready and open to stakeholder involvement.* School leaders play a key role in empowering staff to be involved and open to parents and members of the local community.

Box 1.3. Multiple accountability: Lessons from corporate governance

Within the field of corporate governance, some countries have also moved to systems of multiple accountability. In the United States and the United Kingdom for example, so called "Say-on-Pay" regulations have enabled shareholders to express their voice by voting on the pay policy of the company's executive officers. This vote does not focus on pay itself, but rather on the balance between compensation and performance of the corporation. Proposals that pass the majority threshold are not necessarily binding for the executive board. However, they do exert pressure on the board members to reflect on executive pay and its efficacy to deliver performance. Recent research has shown that Say-on-Pay appears to lead to large increases in market value, profitability and long-term performance in large corporations (Cuñat, Gine and Guadalupe, 2012).

Interestingly, these voluntary initiatives are similar to the Swiss proposal known as the "Minder-Initiative", (named after the entrepreneur Thomas Minder) which was approved by referendum on 3 March 2013. The implementation of the Minder-Initiative requires the remuneration system of stock-traded enterprises to be more transparent and the shareholders' vote on the remuneration system of the companies' boards and executives is binding. However, the boards are free to decide on the modalities of the vote, potentially circumventing the idea behind the initiative.[2]

Capacity building

As education systems must increasingly respond to new societal, economic and individual needs, it is arguably the local level that is most challenged by these developments. A key element of successful policy reform implementation is ensuring that local stakeholders have sufficient capacity to meet this challenge. In particular, they need

adequate knowledge of educational policy goals and consequences, the ownership and willingness to make the change, and the tools to implement the reform as planned. Without these, the best policy reform risks being derailed at the level where it counts most: the classroom. It is at this level that education policies must be implemented, and it is here that they either succeed or fail.

In very simple terms, *capacity building* for governance can be described as the process of helping all actors to acquire and use information relevant to successful policy implementation. Access to this information and understanding how to use the information are defined as "knowledge" (Fazekas and Burns, 2011; Hess and Ostrom, 2007). Capacity building strives to provide different actors with effective and efficient ways to access and use knowledge in local educational contexts in order to achieve desired outcomes.

Target groups for capacity building can be divided into individual, institutional and societal levels, all of which are strongly interrelated (United Nations Economic and Social Council, 2006). In education and the public services, the definition can be extended to include the system level. In this case, capacity building is defined as follows for each of the different levels:

- **Individual level:** Finding ways to support individuals (parents, teachers, headmasters and local policy makers) as they face the demands of new developments in the local context by building on existing knowledge (human resources and knowledge management).
- **Institutional level:** Supporting existing institutions in forming policies, effective organisational structures and good management (this includes building learning organisations) (OECD, 2015b).
- **System level:** Finding efficient ways to support system level actors (e.g. policy makers, teacher unions) to be able to fulfil their roles in designing, implementing and evaluating educational policies.
- **Societal level:** Striving towards more interactive and responsive public administration, and also working to forge links between public sectors to improve the quality and efficiency of governmental service delivery.

Capacity building takes place on two dimensions: *vertically*, through interventions from other levels (for example, from central government to local administration). It is important to recognise that this is a dynamic process and that capacity building in both directions (i.e. from the central and regional levels to local level as well as from the local level to the regional and central levels) is important for efficient education governance. Capacity building can also take place across a particular level with different stakeholders, i.e. *horizontally*. Horizontal capacity building involves sharing experiences and knowledge of efficient ways of implementing policies into practice and also sharing outcomes of the implementation.

Key elements in both an individual's and an institution's capacity building are:

- access to information and the ownership to be willing to use that information
- the ability and tools required to make a change efficiently and as intended, and
- reinforcing desired changes in behaviour to build new reflexes and new patterns of working.

Capacity building needs will be different for policy makers, school leaders, teachers, and parents. In education, there is an on-going challenge in many systems for capacity building for evidence-informed policy making and practice, empowering school leaders for accountability and also responsibility, and redefining the roles of teachers as education professionals.

The use of data

One often over-looked area is the capacity to handle data, both for local government and in schools. Masses of data are available through assessment and monitoring systems, indicators of effectiveness, targets, inspection, and review programs. Methods for accessing information and, consequently, analysing and interpreting it are not self-evident. This is not a new problem: as early as the 1970s, it was observed that much of the relevant data were not available for schools or at least not in a form which could be easily used (Levin, 1974).

However the increased prevalence of data (from student exams, school and teacher evaluations, and a host of other sources) can significantly alter accountability structures in education. Although designed to increase transparency and accountability of education systems, there is a large body of research on the various ways this process can be disrupted or not work as intended. Schildkamp et al. (2014) identify three discrete categories: *non-use, misuse*, and *abuse:*

1. *Non-use:* data is not collected or capacity is lacking to allow for its use. This also includes actors choosing not to use data that is contrary to their argument or beliefs.
2. *Misuse:* data is poorly collected (quality concerns), incorrectly interpreted (analysis or capacity issues) or does not provide adequate answers to be useful for decision-making.
3. *Abuse:* sample or data are manipulated to yield more favourable results, or results in unintended consequences (for example, narrowing the curriculum to improve student scores on tested subjects).

These are serious issues. Appropriate use of data for decision-making requires that local administrators and educators themselves become experts in interpreting data and transforming it into knowledge. This also requires a governance structure that allows for proper circulation and collection of data and provides the correct incentives for its use. Yet all too often this is not the case: Blanchenay et al. (2014) provide an example of local governance decisions being taken on the basis of traditional mechanisms and funding streams rather than the set of (readily available) data.

While many of these arguments focus on the issue from an efficiency perspective (i.e. better use of data enables better and more efficient decision-making), there is also an equity element at play. In many OECD countries wealthier districts or municipalities are more likely than smaller or poorer districts to fully use available data, often due to capacity issues in the ability to analyse and interpret such data. Similarly, upper and middle-class parents are more likely to use school achievement data to place their child in the best-performing schools (see Blanchenay and Burns, this volume, for a more detailed discussion). Parents with lower incomes (including, in many countries, high proportions of immigrant parents) may often lack the capacity to use such data, or indeed base their decisions on other factors, such as geographical proximity and the availability of public transport to access the school (Elacqua et al., 2006).

Both the efficiency and equity arguments are important and suggest that the use of data for decision-making is one of the key needs of a modern education system. This raises important questions from a governance perspective:

- What type of data should be collected (in particular, what balance between qualitative and quantitative data)? At which level? By whom? And what for?
- How well does access to data enable better accountability, with more carefully crafted incentives and responsibilities better tailored to local context?
- What is the best way to create capacity for the use of data (among local decision-makers and central authorities, as well as school administrators and teachers)?
- Is it possible to have "too much" data?

Strategic thinking

Modern governance increasingly relies on strategic thinking to balance the immediate needs and urgencies with longer-term planning and steering of the system. This is not just an education issue, but rather one that touches on all public sectors. Earlier in this chapter the case was made for the concept of a *smarter state* (see Box 1.1) or a strategic state, that "targets its investments to maximize growth in the face of hard budget constraints [and] departs both, from the Keynesian view of a state sustaining growth through demand-driven policies, and from the neoliberal view of a minimal state confined to its regalian functions (public order, basic services)" (Aghion, 2012). In times of economic and fiscal constraint, the argument is that we can no longer afford business as usual. A strategic state implies building the systemic capacity of the government to improve policy design, steering, and implementation over the long-term.

This is a challenge in education, as in many other public sectors. Although it is often argued that increasing decentralisation and increasing school autonomy have contributed to reducing the time available for strategic and system thinking, in the sense that the time required to manage the day-to-day of a more complex system takes away from longer-term thinking, it is clear that this is not the only issue. Difficulties in reconciling time spent on strategy and the ability to design and, crucially, deliver on a long-term agenda are due at least as much to the requirements of the political timeline for voting and agenda setting (OECD, 2009; Blanchenay and Burns, this volume). Regardless of where the problem stems from, there are serious problems with the capacity to engage in and deliver on strategic thinking in many countries, especially outside larger cities.

In general, the central level plays a crucial role in supporting strategic thinking at the local level: capacity building, providing information and offering frameworks. A number of countries have experimented with techniques of strategic thinking in order to find consensus on mid-term national strategies, for example through open consultations and in building and designing visions for the future. The processes are important but very complicated to run, especially given the speed of change and expectations for quick government responses to demands and events. Yet strategic thinking is more and more necessary in complex systems, which require both a holistic vision and the flexibility to deal with change. As one country representative remarked at an OECD conference on this issue: "it is no longer enough to write a white paper and say we are done with the topic".

In order to enhance a system's ability for long-term policy design, some basic preconditions need to be addressed. These include the integration of different types of knowledge and the enhancement of trust between different actors. There is also a

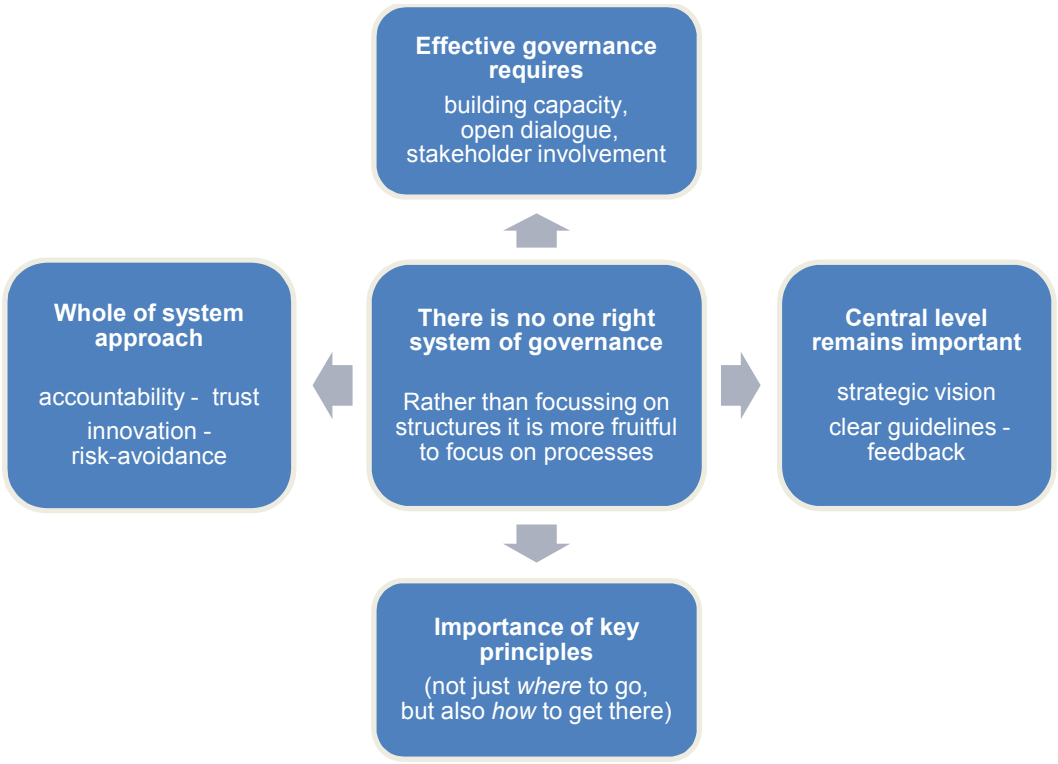
fundamental question of how to design and maintain a long-term strategy in the face of new forces in education such as the 24-hour-media news cycle and school rankings, which are easily picked up and utilised to push for quick changes.

Principles of modern governance

These observations and the work of the GCES project in CERI/OECD have generated five key principles of modern governance:

1. There is no one right system of governance. Almost all governance structures can be successful in education under the right conditions. Successful systems range from fully centralised to almost completely decentralised; some delegate great autonomy to lower levels; in others the central level holds the key to crucial decisions. The number of levels, and the power at each level, is not what makes or breaks a good system. Rather, it is the strength of the alignment, the involvement of actors, and the processes involved in governance and reform. While structures take up a lot of space in the discussion about successful governance, it is more fruitful to focus on processes.
2. A whole of system approach is essential. Education systems must resolve tensions between potentially conflicting forces such as accountability and trust, innovation and risk-avoidance, and consensus building and making difficult choices. Finding the right balance (or, perhaps more accurately, the right combination of mutually reinforcing dynamics that are designed to strengthen both accountability and trust, for example), will depend on the context and history of the system, as well as the ambitions and expectations for its future. A whole of systems approach also works to align roles and responsibilities across the system, improving efficiency as well as reducing potential overlap or conflict.
3. Effective governance works through building capacity, open dialogue, and stakeholder involvement. However it is not rudderless: involvement of a broader range of stakeholders only works when there is a strategic vision and set of processes to harness their ideas and input.
4. Even in decentralised systems, the national or state level remains very important in triggering and steering education reform. The central level most often provides the system-wide vision needed to enable effective delivery of reform as well as equitable access and outcomes for students. It can also be instrumental in developing clear guidelines and goals, and providing feedback on the progress on those goals, the building blocks of any successful governance and reform process.
5. There is a need to develop key principles for system governance (not just agreement on where to go, but how to get there). The key principles must be built on whole of system thinking and work to align the different actors and levels. Examples of goals include reducing the drop-out rate and improving student attainment. Examples of key principles underlying the governance and decision-making used to achieve those goals would be having a system that is open, inclusive, positive, and evidence-informed.

Figure 1.3. Principles of modern governance



This volume

The volume is organised in four parts, tied closely to the work done by the OECD/CERI Governing Complex Education Systems project.

Part I: Setting the stage: Modern Governance Challenges in Education

Following this introduction by the OECD Secretariat, **Mark Mason's** *Complexity Theory and Systemic Chang in Education Governance* continues setting the stage of this publication. The chapter focuses the notion of complexity and what it means for education governance. Departing from an overview of general complexity theory, Mason discusses the concept's relevance for education governance and how the lens of complexity theory can aid policy making.

Henno Theisens' chapter problematizes the growing complexity of society and its demands on governance. The chapter *Education Governance: Hierarchies, Networks and Improvisation* argues that a return to centralized planning is neither feasible nor desirable: central rational planning has become too unwieldy for the dynamic and fast paced challenges of modern societies. While the ubiquitous trend towards decentralisation and marketization has serious shortcomings, it has succeeded in moving decision-making processes closer to the respective stakeholders. Theisens proposes an approach labelled "governance through networks" where informal and dynamic networks take the place of central planning and marketisation. The civil servant's role is conceptualised as that of an actor within the networks rather than that of a representative of a hierarchical government and rational planning. Nevertheless the central level takes a steering and enabling role, providing the overall framework in which the networks function.

Part II – The Role of Accountability in Governing Complex Systems

The second part of this volume revolves around the role and consequences of accountability mechanisms for governance and those who are held accountable. Decentralisation has been accompanied in many countries with a greater use of assessment and evaluation systems aimed at holding lower levels accountable for their practices. This pertains to lower levels of governance as well as ultimately to schools and teachers. **William Smith** examines school accountability systems based on student test scores in *Exploring Accountability: National Testing Policies and Student Achievement.* The chapter finds that high-stakes systems have adverse effects on teaching practices such as teaching to the test and narrowing the curriculum.

Edith Hooge's *Making Multiple School Accountability Work* discusses the inclusion of local stakeholders to improve accountability on the horizontal level. A system of horizontal accountability draws on insights of local stakeholders in areas such as priority setting and performance evaluation and uses this in combination with student testing outcomes to determine school and regional performance. If done correctly, such horizontal accountability mechanisms give schools the means to present a fuller picture of their performance to central governance levels. On another level, building capacity of schools to accommodate voices of local stakeholders can satisfy demands of transparency and involvement and facilitate the acceptance of education reforms among the community.

Part III – Capacity and the Use of Knowledge

Lorenz Lassnigg's chapter reminds us that policy does not operate in a vacuum. Taking Austria's complex multilevel system as example, *Complexity in a Bureaucratic-Federalist Education System* explores a number of issues related to the alignment of different logics present in structurally complex systems. Lassnigg describes how Austria's governance structure creates a tight corset of power distribution intended to increase political representation. The chapter discusses that in doing so the system leads to a large gap between formal structures and informal practices, paradoxically exacerbating the unpredictability it seeks to reduce.

Lassnigg proposes a more active role of practitioners in policy research, with a more network-oriented, collaborative role of local actors in governance. However, the chapter makes clear that political power relations and politics' normative beliefs can be hard to change and that timing and adaptation of policies to country contexts are crucial to effect change.

The second chapter in this area focuses directly on capacity and the use of knowledge in education. Based on the example of England (United Kingdom), **Philippa Cordingley** discusses how teacher involvement in research can be used to build teachers' capacity in their own research to improve instruction. In her chapter *Knowledge and Research Use in Local Capacity Building,* Cordingley distinguishes between teachers' engagement with external research and, in a more advanced state, teachers' engaging in their own research with the goal to improve instruction – not only for local gain but also to contribute to improved instruction practices which can be scaled up to the system level.

The author emphasises the responsibility of governance to facilitate teachers' engaging with research not least by making research tools available that are practical in the specific work environment in schools. Teachers' role then ultimately is to identify so-

called "wicked issues" that are taken up in local teacher research and external research alike to improve instruction.

Part IV – Complexity in Policy Making: Thinking Strategically

The volume's final part closes the loop and returns to complexity theory. The chapters in this section use complexity theory as lens for policy makers to facilitate successful reform. **Patrick Blanchenay and Tracey Burns'** chapter *Policy Experimentation in a Complex Environment* discusses the consequences for policy making when acknowledging the complexity of systems. The chapter discusses the profound dilemma between focused experiments and scaling up to the larger network of stakeholders given that complex systems are characterized by unpredictability, where a small difference in context can lead to fundamentally different results. In order to evaluate experiments on a larger scale without implementing the respective policy in the whole system, the chapter proposes the concept of ecosystem experimentation. Ecosystems are conceptualized as networks of actors that are to a reasonable degree self-contained. Identifying networks with only weak links to other networks, ecosystem experimentation strikes a balance between the complexity of the system and its unpredictable consequences and evaluating its effect in reasonable diverse and large network of actors.

Lex Borghans, Trudie Schils and Inge de Wolf examine the Netherland's experience with policy experimentation in the chapter *Experimentalism in Dutch Education Policy*. The chapter explores the scope of experimentation and related innovation in the Dutch education system. It describes examples of the various forms of experiments carried out as well as dilemmas and lessons learned related to experimentation. The authors observe that the involving and supporting education practitioners, ensuring schools' capacity as well as knowledge dissemination are critical for successful experimentation.

Inherent to policies and reforms is the ability to take risks and innovate. However taking risks can (and often does) result in failure, which is difficult to reconcile with standard accountability mechanisms and political imperatives. In *Learning to Fail, Not Failing to Learn,* **Tracey Burns and Patrick Blanchenay** discuss the need to think about risk, not just as something to be contained, but as an intrinsic part of innovation and change to improve systems. As such, controlled risk-taking needs to be better governed and accepted as part of policy-making and implementation, for example in the use of experimentation. Failure could thus be integrated into system functioning and used as an opportunity to learn. This challenging notion requires a change from using failure to assign blame, or reinterpreting failure as a success. The kind of system change will only be possible if the system is designed to recognise failure as an inherent part of reform and experimentation.

Concluding note

The search for new modes of governance that allow policy makers to address 21st century education governance challenges will certainly continue in the years to come. In a decade from now we may still be noting the same challenges in balancing accountability and innovation, and finding consensus and making difficult decisions. The agenda that has been set out here, and the challenge to create the open, adaptable, and flexible governance systems necessary for governing complex systems, is not an easy one. Pressures will continue to mount in terms of expectations for participatory

governance as our world continues to become more networked and the role of social media – and media more broadly – emerges as a key actor. The rise in power of these actors will likely recast the processes (and potentially structures) of prioritising, steering and accountability.

From our perspective the greatest challenge of current educational governance is creating a strong and constructive accountability system that balances the monitoring and control required to ensure efficient system functioning with a push for system improvement and support for the broader holistic goals of education. What elements make up such an accountability system? And how can this be achieved in the particular contexts and traditions of each education system, especially given the tightening grip of finance ministries on education spending?

One last note. The discussion outlined in this chapter has a common theme running through it that has not been explicitly developed, that of trust. Trust is essential to good governance across a variety of policy areas, including education. It is essential for the functioning of our systems, for the ownership and implementation of policies and reforms, and for basic collaboration and teamwork. Yet we know relatively little about how trust is developed and sustained over time, or restored if broken (Cerna, 2014). We will thus return to the theme of trust in the final concluding chapter, to examine this seemingly simple, yet decidedly complex topic more thoroughly.

Notes

1. www.oecd.org/edu/ceri/gces.
2. For further details see www.ethosfund.ch/e/news-publications/news.asp?code=303 (English) and accompanying report (in French, German).

References

Aghion, P. (2012), *Growth Policy and the State Implications for the Design of a European Growth Package*, Growth Commission, Institute of Government, London School of Economics.

Blanchenay, P., T. Burns and F. Köster (2014), "Shifting responsibilities - 20 years of education devolution in Sweden: A Governing Complex Education Systems case study", *OECD Education Working Papers*, No. 104, OECD Publishing, Paris, http://dx.doi.org/10.1787/5jz2jg1rqrd7-en.

Borman, G.D. et al. (2003), "Comprehensive school reform and achievement: A meta-analysis", *Review of Educational Research*, Vol. 73/2, pp. 125-230.

Busemeyer, M. and J. Vossiek (2015), "Reforming education governance through local capacity-building: A case study of the 'Learning Locally' Programme in Germany", *OECD Education Working Papers*, No. 113, OECD Publishing, Paris, http://dx.doi.org/10.1787/5js6bhl2mxjg-en.

Cerna, L. (2014), "Trust: what it is and why it matters for governance and education", *OECD Education Working Papers*, No. 108, OECD Publishing, Paris, http://dx.doi.org/10.1787/5jxswcg0t6wl-en.

Dror, Y. (1986), *Policy Making Under Adversity*, Transaction Press, New York.

Elacqua, G., M. Schneider and J. Buckley (2006), "School choice in Chile: Is it class or the classroom?", *Journal of Policy Analysis and Management*, Vol. 25/3, pp. 577-601.

Fazekas, M. and T. Burns (2012), "Exploring the Complex Interaction Between Governance and Knowledge in Education", *OECD Education Working Papers*, No. 67, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k9flcx2l340-en.

Hooge, E., T. Burns and H. Wilkoszewski (2012), "Looking beyond the numbers: Stakeholders and multiple school accountability", *OECD Education Working Papers*, No. 85, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k91dl7ct6q6-en.

Hopfenbeck, T. et al. (2013), "Balancing trust and accountability? The assessment for learning programme in Norway: A Governing Complex Education Systems case study", *OECD Education Working Papers*, No. 97, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k3txnpqlsnn-en.

Mazurkiewicz, G., B. Walczak and M. Jewdokimow (2014), "Implementation of a New School Supervision System in Poland", *OECD Education Working Papers*, No. 111, OECD Publishing, Paris, http://dx.doi.org/10.1787/5jxrlxrxgc6b-en.

OECD (2015a), "Final NAEC synthesis – New approaches to Economic Challenges", OECD, Paris, www.oecd.org/mcm/documents/Final-NAEC-Synthesis-Report-CMIN2015-2.pdf.

OECD (2015b), *Schooling Redesigned: Towards Innovative Learning Systems*, Educational Research and Innovation, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264245914-en.

OECD (2013), *Synergies for Better Learning - An International Perspective on Evaluation and Assessment*, OECD Reviews of Evaluation and Assessment in Education, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264190658-en.

OECD (2009), *Working Out Change: Systemic Innovation in Vocational Education and Training*, Educational Research and Innovation, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264075924-en.

Sabelli, N.H. (2006), "Complexity, technology, science, and education", *The Journal of the Learning Sciences*, Vol. 15/1, pp. 5-9.

Schildkamp, K., L. Karbautzki and J. Vanhoof (2014), "Exploring data use practices around Europe: Identifying enablers and barriers", *Studies in Educational Evaluation*, Vol. 42, pp. 15-24.

Snyder, S. (2013), "The simple, the complicated, and the complex: Educational reform through the lens of complexity theory", *OECD Education Working Papers*, No. 96, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k3txnpt1lnr-en.

United Nations Economic and Social Council (2006), "Definition of basic concepts and terminologies in governance and public administration", Committee of Experts on Public Administration, Agenda item 5, *Compendium of Basic Terminology in Governance and Public Administration*, E/C.16/2006/4, pp. 1-15.

Van Twist, M. et al. (2013), "Coping with very weak primary schools: Towards smart interventions in Dutch education policy", *OECD Education Working Papers*, No. 98, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k3txnpnhld7-en.

Waslander, S., C. Pater and M. van der Weide (2010), "Markets in education: An analytical review of empirical research on market mechanisms in education", *OECD Education Working Papers*, No. 52, OECD Publishing, Paris, http://dx.doi.org/10.1787/5km4pskmkr27-en.

Wilkoszewski, H. and E. Sundby (2014), "Steering from the centre: New modes of governance in multi-level education systems", *OECD Education Working Papers*, No. 109, OECD Publishing, Paris, http://dx.doi.org/10.1787/5jxswcfs4s5g-en.

Annex 1.A1: Central outputs of the GCES project

Case study series

- *Flanders, Belgium:* Research on this case study began in early 2016. The case study will examine the implementation of the core curriculum, and in particular the process for defining attainment targets and developmental goals. Although it has been in existence for quite some time, the question arises: how well has the core curriculum been implemented in practice?
- *Reforming Education Governance through Local Capacity-Building:* A Case Study of the "Learning Locally" Programme in Germany (Busemeyer and Vossiek, 2015) finds that the "Learning Locally" Programme can be regarded as a success due to the fact that it had a lasting and probably sustainable impact. It reveals that a number of local factors influence the relative effectiveness of the implementation of the programme.
- *Implementation of a New School Supervision System in Poland* (Mazurkiewicz, Walczak and Jewdokimow, 2014) explores the strategies, processes and outcomes of an education reform in Poland which was introduced in 2009 and substantively changed the school inspection system.
- *Shifting Responsibilities:* 20 years of Education Devolution in Sweden (Blanchenay et al., 2014) examines the consequences of important education decentralisation reforms that took place in Sweden since the early 1990s.
- *Coping With Very Weak Primary Schools:* Towards Smart Interventions in Dutch Education Policy (van Twist et al., 2013) looks at the effectiveness of policy instruments aimed at reducing the number of underperforming primary schools in a system with a long tradition of school autonomy.
- *Balancing Trust and Accountability*? The Assessment for Learning Programme in Norway (Hopfenbeck et al., 2013) explores the implementation strategies used Norway to enhance formative assessment in its schools.

Working paper series

- *The Educational Roots of Trust* (Borgonovi and Burns, 2015) examines the association between education and levels of interpersonal trust, using data from the Survey of Adult Skills, a product of the OECD Programme for the International Assessment of Adult Competencies (PIAAC).
- *Steering from the Centre: New modes of Governance in Multi-level Education Systems* (Wilkoszewski and Sundby, 2014) explores innovative governance strategies for the central level in education systems. It identifies core features of multilevel governance and introduces a basic analytical categorisation of modes of governance.

- *Trust: What It is and Why it matters for Education and Governance* (Cerna, 2014) analyses the centrality of trust for policymaking and current governance issues. Trust enables stakeholders to take risks, facilitates interactions and co-operation, and reduces the need for control and monitoring.
- *The Simple, the Complicated, and the Complex: Educational Reform through the Lens of Complexity Theory* (Snyder, 2013) explores complexity theory and its applications for educational reform. After discussing the key concepts of complex adaptive systems, the paper defines the differences between simple, complicated, and complex approaches to educational reform.
- *Exploring the Complex Interaction Between Governance and Knowledge in Education* (Fazekas and Burns, 2012) asks the question of how governance and knowledge mutually constitute and impact each other in complex education systems.
- *Looking Beyond the Numbers: Stakeholders and Multiple School Accountability* (Hooge, Burns and Wilkoszewski, 2012) analyses trends in accountability mechanisms and processes and argues that regulatory and school performance accountability can be usefully augmented by involving multiple stakeholders.

Conferences

- *Trust in Education* (7 December 2015 in The Hague, the Netherlands) focused on building and sustaining trust in education. It brought together state of the art research with country examples of the role of trust in education, with a focus on accountability, professionalism, and responsibility.
- *The Use of Data in Educational Governance* (12-13 February 2015 in Tallinn, Estonia) focused on the use of data for education governance. The main themes included the challenges of the use of data in education, some strategies that have been applied to tackle these challenges, and the kinds of support needed at different governance levels.
- *Understanding Complexity:* The Future of Education Governance (10 February 2014 in Oslo, Norway) revolved around the impact of complexity on education governance. At the conference were discussed the challenges of complexity for education, some of the approaches to cope with these challenges, as well as the identification of gaps in our knowledge base.
- *Effective Multilevel Governance in Education* (17-18 June 2013 in Paris, France) focused on two main themes in effective multilevel governance: transparency and trust. The conference was a joint collaboration between the OECD's Centre for Educational Research and Innovation (CERI) and UNESCO.
- *Effective Governance on the Local Level* (15-17 April 2012 in Warsaw, Poland) looked at the role of local stakeholders in the governance of complex education systems. The conference asked about the place of local authorities and schools in the governance process, how local authorities and schools can be ensured to have the capacity to govern their local systems and how local stakeholders can hold local authorities accountable.

- *Effective Governance from the Centre* (21-22 November 2011 in The Hague, the Netherlands) focused on the role of central government in complex, multilevel systems of governance. Even as regional, local and school levels receive more autonomy, the role of the centre is still crucial as it is being held accountable for education outcomes and is in the best position to ensure a common direction and set priorities.
- *The GCES Launch Conference* (28-29 March 2011 in Oslo, Norway) contributed to defining the scope and direction of the project. The conference explored which governance mechanisms and knowledge options facilitate effective steering of complex education systems by bringing together an international group of senior policy makers and researchers.

Chapter 2.

Complexity theory and systemic change in education governance

Mark Mason

Hong Kong (China) Institute of Education

Education governance has among its principal responsibilities initiating and sustaining positive change – whether at system, district or school level. The insights offered by complexity theory suggest a radical rethinking of some of the more traditional notions about how this might be achieved. This paper accordingly considers the challenge of sustainable change in education from the perspective of complexity theory. Complexity theory's concept of emergence implies that, given a significant degree of complexity in a particular environment – whether an education system or a particular school – new properties and behaviours emerge that are not necessarily contained in the essence of the constituent elements, or easily able to be predicted from a knowledge of initial conditions. These concepts of emergent phenomena form a critical mass, associated with notions of lock-in, path dependence, and inertial momentum, contribute to a perspective on continuity and change that indicates what conditions might need to be in place for the emergence of sustainable, positive, system-wide change in education.

Introduction

Some of the most interesting and worthwhile implications of complexity theory include the insights it offers into understanding change – and, by implication, continuity. Perhaps more realistically than a *theory* of change in the social sciences, it offers a metaphor, or a lens, through which we might better understand what it takes to initiate and to sustain systemic change. The conceptualisation of change in complexity theory provides some insight into what manner of intervention stands the most chance of being sustained – a question of considerable importance for education governance.

In this chapter, then, I consider the challenge of sustainable change in education from the perspective of complexity theory. To do so I will firstly describe the background and the core concepts of complexity theory. Building on these concepts, I discuss how change comes about in complex systems and how policy making can be rethought to make use of the logics underlying complexity and initiate sustained change in complex education systems. It is suggested that the complexity generated by a network of multiple integrated and mutually supportive initiatives will precipitate and sustain change more successfully than will individual and isolated initiatives. Initiating and sustaining change in a complex system thus requires sensitivity to a multiplicity of factors that compound and mitigate each other in recursive and cyclical patterns, and the design of an integrated suite of interventions on multiple levels, from multiple points, and that take into account this multiplicity of factors.

Consider two frequently asked questions: about the origins of life itself; and about how consciousness emerges from an agglomeration of biological cells – simple questions that should get us off to an easy start. What we can at least say is that life, and indeed consciousness, are best understood as *emergent phenomena*: while the brain is a complex arrangement of billions of neurons functioning according to the laws of cell biology, the phenomenon of mind emerges as much more than a biological agglomeration of nerve cells. The principle of emergent phenomena on account of increasingly complex networks among constituent elements has been used by the theoretical biologist, Stuart Kauffman (1992), to explain the origins of life. As the Nobel laureate physicist, Phil Anderson (1972, cited by Waldrop, 1993: 82), has argued, "At each level of complexity, entirely new properties appear. [And] at each stage, entirely new laws, concepts, and generalisations are necessary…. Psychology is not applied biology, nor is biology applied chemistry."

It is important to note at the outset that the notions of scale and complexity are what underlie this principle of *emergence*. New properties or behaviours emerge when sufficient numbers and varieties of constituent elements or agents cluster together to form a sufficiently complex and dynamic arrangement of incredible scale. The concept of emergence implies that, given a significant degree of complexity in a particular environment, new properties and behaviours emerge that are not necessarily contained in the essence of the constituent elements, or easily able to be predicted from a knowledge of initial conditions. These concepts of emergent phenomena from a critical mass, associated with notions – that we will come across shortly – of lock-in, path dependence, the "economics of increasing returns", and inertial momentum, contribute to a perspective on continuity and change that indicates what conditions might need to be in place for the emergence of sustainable, positive, system-wide change and development in education.

To reiterate, complexity theory is, first and last, about reaching critical mass among the diverse range of factors, elements and agents that constitute a particular environment.

It offers, in other words, a dynamic and system-wide perspective on how sustainable change, characterised by new properties and behaviours in the education system, emerges from the interaction of a myriad factors in the economic, political, social and cultural environments in which education is situated. Other theories of change have sought "the levers of history" – a metaphor I am going to suggest is inappropriate – in economic structures, in human agency, and in combinations of these and other factors that include or exclude either or both. Complexity theory offers a theory of change that might be said to encompass all of these and more, and that might offer the most helpful insight yet into how educational development and change might be rendered sustainable.

Complexity theory: Interaction and adaptation dynamics creating emergent behaviours

Developed in the fields of physics, biology, chemistry and economics, complexity theory arises in some senses out of chaos theory in that it shares chaos theory's focus on the sensitivity of phenomena to initial conditions that may result in unexpected and apparently random subsequent properties and behaviours. Chaos theory suggests that even a very slight degree of uncertainty about initial conditions can grow inexorably and cause substantial fluctuations in the behaviour of a particular phenomenon – Edward Lorenz's "butterfly effect" (Lorenz, 1963). Perhaps more importantly, complexity theory shares chaos theory's concern with wholes, with larger systems or environments and the relationships among their constituent elements or agents, as opposed to the often reductionist concerns of mainstream science with the essence of the "ultimate particle".[1]

Complexity theory concerns itself with environments, organisations, or systems that are complex in the sense that very large numbers of constituent elements or agents are connected to and interacting with each other in many different ways. These constituent elements or agents might be atoms, molecules, neurons, human agents, institutions, corporations, etc. Whatever the nature of these constituents, the system is characterised, as Waldrop (1993) has described, by a continual organisation and re-organisation of and by these constituents:

> …into larger structures through the clash of mutual accommodation and mutual rivalry. Thus, molecules would form cells, neurons would form brains, species would form ecosystems, consumers and corporations would form economies, and so on. At each level, new emergent structures would form and engage in new emergent behaviours. Complexity, in other words, [is] really a science of emergence. (Waldrop, 1993: 88)

One of the most important insights of complexity theory is this notion of emergence which implies that, given a sufficient degree of complexity in a particular environment, new (and to some extent unexpected) properties and behaviours emerge in that environment. The whole becomes, in a very real sense, more than the sum of its parts in that the emergent properties and behaviours are not necessarily contained in or easily able to be predicted from the essence of the constituent elements or agents. A central concern of complexity theory is thus on the relationships among the elements or agents that constitute a particular and sufficiently complex and dynamic environment or system. Once a system reaches a certain critical level of complexity, otherwise known as the critical mass, a phase transition takes place, which makes possible the emergence of new properties and behaviours and a new direction of self-sustaining momentum. A certain critical level of diversity and complexity must be reached for, say, an education system to

achieve this sustainable autocatalytic state – that is, for it to maintain its own momentum in a particular direction. The model posits the phase transition as a fundamental law of increasing complexity, but the specific details of this phase transition – when and how it occurs, what properties and behaviours emerge – are contingent on specific contextual factors and are probably unique to that particular context.

Complexity theory makes no claim to predict what is essential and what can be marginalised in the search for "the levers of history". In this sense, and as will become clearer, the perspective of complexity theory on change shows how inappropriate this analogy is, with its connotations of single, powerful causes. As a research paradigm, complexity theory cautions us not to marginalise or dispense with what is apparently trivial or inexplicable. What may appear to be marginal may be part of the complexity of a system, and may be constituent of the critical level above which emergent properties and behaviours become possible.

Complexity theory suggests that it is in the dynamic interactions and adaptive orientation of a system that new phenomena, new properties and behaviours, emerge, that new patterns are developed and old ones change. It seeks the sources of and reasons for change in the dynamic complexity of interactions among elements or agents that constitute a particular environment. It is in this sense that seemingly trivial accidents of history may increase dramatically in significance when their interactions with other apparently minute events combine to produce significant redirections in the course of history, significant shifts in the prevailing balance of power. Complexity theory can accept the existence of certain essential generative elements in a particular field, but suggests that the field as a whole is much more than merely predictably determined by the primary generative element. While this may be a trigger, and indeed only one of many triggers, of subsequent phenomenal developments, complexity theory suggests that it is the manifold interactions among constitutive elements or agents that are responsible for the phenomena, patterns, properties and behaviours that characterise a particular field.

The successive addition of new elements or agents to a particular system multiplies exponentially the number of connections or potential interactions among those elements or agents, and hence the number of possible outcomes. This is an important attribute of complexity theory, in that the connections among individual agents or elements assume an importance that is critical to complexity theory's assertions about emergent properties. This emergence becomes possible by virtue of the exponential relationship between the elements or agents and the connections among them. The essence of the individual elements or agents that constitute a particular system does not alone provide the key to understanding that system. Complexity theory draws attention to the emergent properties and behaviours that result not only from the essence of constitutive elements, but more importantly, from the connections among them. The focus thus shifts from a concern with decontextualised and universalised essence to a concern with contextualised and contingent complex wholes. Complexity theory echoes Foucault's emphasis on "polymorphous correlations in place of simple or complex causality" (cited in Harvey, 1990: 9). Admittedly, complexity theory does suggest that new properties and behaviours will emerge out of these "polymorphous correlations", but the point is that the possibility is lessened of an accurately predicted causal relationship from known initial conditions to these emergent phenomena.

The roles of lock-in, inertial momentum and the snowball effect in changing the status quo

In the light of complexity theory, I would define power, or, more simply, the prevailing balance of the *status quo*, of the way things are, as the directional course of the phenomenon that enjoys the dominant inertial momentum[2] over other competing phenomena. The prevailing *status quo* will sustain, and indeed increase, its dominance by virtue of what can be simply and analogously understood as the snowball effect. Individual and apparently trivial accidents in the purview of the dominant structure's momentum will be gathered up in its path; those outside of its purview will remain marginal and ineffective unless and until sufficient momentum in a different direction is sustained by sufficient complexity of a different, if related, concatenation, or network – to use a different metaphor – of originally trivial events. How radical the power shift is will depend on the degree of difference in strength and direction – as in velocity or in vector analysis, but rather more amorphously – between the existing and the emerging power structures. The term path-dependence, allied to the notion of lock-in, illuminates this idea by suggesting that the inertial momentum of a particular phenomenon will sustain its direction and speed along a particular path, that a phenomenon is describable in terms of the direction of its path, and that it will continue in that path to the point where sufficient inertial momentum of a competing phenomenon results in a redirection of that path. In this manner, or, analogously, good educational institutions or systems will sustain and probably increase their own momentum, and weaker educational institutions or systems will likewise compound the failure of their students, thereby further weakening themselves in an endless and vicious cycle.

Box 2.1. Vicious and virtuous circles in the Dutch education system: coping with very weak primary schools

The Dutch school system is consistently ranked as one of the systems with the highest levels of school autonomy in the OECD (e.g. OECD, 2011a). Beginning in the late 1980s, the Dutch education system increased the autonomy of its schools, giving them almost complete authority to govern themselves. The Dutch Ministry of Education accordingly relies on a set of indicators to fund schools by "lump-sum" and to monitor educational quality. What happens in schools is the responsibility of schools and – if quality is lacking – of the Inspectorate.

Based on a number of (output) indicators, the Inspectorate assesses the risk of an individual school underperforming. If a school is deemed at risk, it will receive inspection, which will lead to an overall assessment as "normal", "weak" or "very weak". Schools that are rated weak or very weak will receive more intense follow-up inspection, and those that are labelled very weak must improve or be closed down within two years. During these two years, the Inspectorate engages with school boards and monitors the implementation of its recommendations. Alongside this intervention, weak schools are provided with specialised advice and assistance, subsidised by the Ministry and carried out by organisations in the field. Overall, this policy has been successful in reducing the number of schools with negative labels. However, some schools have weakened even further. The results of these interventions are often difficult to predict, mostly because of subtle differences between schools and their contexts that can be neither completely known nor affected at the national level. While different both in terms of their background and in their response to the Inspectorate's label of "very weak", all schools observed share a common element in demonstrating a cyclical dynamic following intervention.

Van Twist et al. (2012) suggest that there is no simple linear flow of cause and effect driving performance upward or downward. Even just the assignment of the label "very weak" can elicit a positive response from one school and a negative response from another, depending on numerous factors that include the local context, history and staffing situation at the school.

Box 2.1. Vicious and virtuous circles in the Dutch education system: coping with very weak primary schools *(cont.)*

In some schools, parents removed their children after it was labelled "very weak", while in others, they seemed to become more involved in the school. A similar pattern was discernible among teachers: in some schools they withdrew and even reacted angrily; in others they started to make changes in the school following the inspection/intervention. In some schools, being stigmatised as "very weak" created division within the staff, or between staff and management; in other schools the same trigger served to bond the team together. Inspectorate intervention thus set in motion a range of possible self-reinforcing cycles. A first wave of exit behaviour by parents might trigger exit behaviour by a wider group of parents; contrasted with this, an initiative of a small group of parents to become further involved might resonate at a larger scale and attract more voluntary engagement. These effects are neither linear nor planned; they are partly unintended, iterative and cyclical processes. Such positive and negative cycles can co-exist within and around a school (i.e. in student, staff, management and parent behaviour). Precipitating and sustaining change in a complex system thus requires sensitivity and thoughtful responses to a multiplicity of factors that compound and mitigate each other in recursive and cyclical patterns.

Source: Van Twist et al., 2013.

The notion of inertial momentum, referring to the snowball effect, or the ever-increasing probability of the development and sustenance of correlated possibilities on account of recently developed phenomena, provides the conceptual link between the principle of emergent phenomena as developed principally in the natural sciences and the notion of socio-historical change in human society. Inertial momentum is, as I have suggested, inextricably related to the phenomenon of power. The power of an existing dispensation or social arrangement to sustain itself and to increase its purview of influence or control is directly related to its inertial momentum, to the aggregate weight of the phenomena of which it is constituted. And this aggregate is the result of the number, scale and diversity of the elements and agents that constitute the social arrangement, and of the degree of complexity of the interactions among them.

This snowball effect can be understood in terms of what the economist Brian Arthur (1989) called "the economics of increasing returns", allied to the idea of "lock-in". A striking example is the dominance of the QWERTY typewriter keyboard (at least in Anglophone countries). When mechanical typewriters were developed, touch-typists had to be slowed down by inefficient keyboard layouts because their increasing dexterity would continually jam the mechanically slow machines. One of the most inefficient designs (by Christopher Scholes in 1873) was the QWERTY layout, which was adopted and mass-produced by Remington. More typists accordingly learned on the QWERTY layout, more companies therefore adopted the same layout, and a virtually unbreakable lock-in of the QWERTY keyboard resulted. Other more efficient keyboard layouts have been designed, but the probability of their breaking the locked-in monopoly of the QWERTY keyboard, particularly now, given the contemporary proliferation of computer keyboards (and, ironically, when we no longer have to worry about the mechanical jamming of the keyboard), is almost zero.

The phenomenon of lock-in is associated with the "spontaneous self-organisation" of systems identified by the Nobel laureate physicist, Ilya Prigogine (1980), in his research on the origins of order and structure at all levels of the universe. The spontaneous dynamics of living systems result from the positive feedback to or self-reinforcement of

phenomena, a process which is characterised by the increased incidence and significance of initially apparently trivial events under the at first random conduciveness of circumstances. While the circumstances in which the positive feedback eventually occurs may have been initially random, the self-reinforcement leading to lock-in of a particular phenomenon reflects an autocatalytic chain of events in the field. The direct implication is of a self-sustaining phenomenon which, while the statistical chances of its appearance may have been negligible at first, emerges adaptively, locks itself in, and sustains its inertial momentum autocatalytically. To complexity theorists, "positive feedback seem[s] to be the *sine qua non* of change, […] of life itself" (Waldrop, 1993: 34). It becomes obvious that Darwin, although he probably never used the term, was a complexity theorist. The process of the emergence of new phenomena and the extinction or adaptation of existing arrangements explains the adaptive orientation and "spontaneous self-organisation" of a system, the "incessant urge of complex systems to organize themselves into patterns" (Waldrop, 1993: 118). Darwin and complexity theory are, in other words, complementary in their explanation of evolution, in their explanation of the nature of change. This autocatalytic sustenance of momentum becomes enormously powerful: any young and idealistic teachers, no matter how energetic, who have gone into weak schools with the intention of turning them around, will report that fighting their momentum is like shouting into the wind. They may touch the lives of a handful of students, but that is probably it. At the risk of stating the obvious, it takes more than the efforts of a few energetic teachers to affect the inertial momentum of a weak school that sustains its weakness autocatalytically.

A salient feature of a theory of increasing returns is that there are, initially at least, multiple possible outcomes. Which outcome is realised in the social sciences is a question of intervention at as many levels as possible: for example, at the macro-structural level and at the intentional human agency level, so that sufficient momentum is generated in a particular direction to displace the inertial momentum of the current dispensation and to create a dominant inertial momentum for the desired changes. The dominant status of a particular social policy, for example, is more a function of that policy's inertial momentum than just the legislation that supports it. Complexity theory renders largely irrelevant the agency-structure debate about which of the two is more important in effecting change. I will consider this debate, whether change can be effected through human agency or whether deeper and more powerful structural forces are at work, in a substantial example in a moment, but it is worth noting here that both structure and agency are important in introducing change that can be sustained, and much else is too.

Working in probability theory, Brian Arthur and others have constructed mathematical models by which it is possible to follow the process of the emergence of one historical outcome, to "see mathematically how different sets of historical accidents could cause radically different outcomes to emerge" (Waldrop, 1993: 46). What this means for successful policy implementation is that positive feedbacks shaped towards a particular outcome need to be created through conscious interventions, so that new patterns are established. Once sufficient momentum is generated in the new (and desired) direction, the positive feedback becomes incorporated into the system autocatalytically, and new phenomena predominantly typical of the desired policy's characteristics, emerge. Changing education systems to rid them of their inequities and inefficiencies will, in other words, require massive interventions at all levels.

Generating momentum to make educational development and change sustainable

In the complexity of the educational environment, the plethora of relevant constituent elements – agents and structures – includes teachers, students, parents, community leaders, the state and its education departments and policies, economic structures and business organisations, NGOs, agencies, and so on. Intervention to differing but sufficient extents in each of these areas is what would probably be necessary to shift a prevailing ethos in education. In other words, change and sustainable development in education, at whatever level, are not so much a consequence of effecting change in one particular factor or variable, no matter how powerful the influence of that factor. It is more a case of generating momentum in a new direction by attention, as I have argued, to as many factors as possible.

Such a conclusion asks a lot more of governments and their education departments, of research analysts and policy-makers, and of donors, aid agencies and development specialists, than has typically been asked in the past. Take the case of a school or an education system where the prevailing ethos is one of failure, where students are, for any number and combination of reasons, not learning. The agency-structure debate invites us to consider whether change can be effected through human agency, or whether deeper and more powerful structural forces are at work. Structuralists, who find in economic factors the primary "levers of history", might suggest that there is little we can do about this as human agents, because the despair that pervades the school and the system is primarily a consequence of the jobless future that awaits school leavers, whether certificated or not. The ethos of the school will not change until the structure and nature of the economy change in such a way as to provide meaningful and worthwhile employment for certificated school leavers. Those on the agency side of the debate might point to the importance of an excellent school leader, or of a committed *corps* of teachers across the system. Complexity theorists would suggest that it is probably both – the structural factors and the influence of human agents – and far more. But because we can never know well enough the combination and salience of factors that are causing the school's or the system's failure, or exactly what it is that will turn things around, our best chance of success lies in addressing the problem from as many angles, levels and perspectives as possible. It's more than that we cannot quantify the salience of any individual factor: we probably cannot even isolate any individual factor's influence in order to assess its salience. This is of course because various factors compound each other's effects in ways that both increase and diminish their aggregate influence.

Box 2.2. The Barefoot College's Solar Night schools programme

The Solar Night Schools Programme is an initiative of the Barefoot College in India's state of Rajasthan that provides access to education for rural children who cannot attend local public schools because they open only during the day – when most children in the region are required by their families to help with subsistence activities centred on agriculture and animal husbandry. In the semi-desert environment, families frequently move with their animals, making it difficult for children who work as herders to adjust to any formal schooling schedule. In most cases the children who attend the Solar Night Schools are poor, low caste and from illiterate families. Most are girls: while their brothers might be sent to "normal" schools where possible, girls are kept at home to help with their families' economic activities. As a consequence, the Solar Night Schools Programme offers the only education option for many girls in the region. To enable the schools to operate at night (when children have finished their household duties) in remote and poor areas where there is no electricity grid, light is provided by solar lanterns made at the Barefoot College.

Box 2.2. The Barefoot College's Solar Night schools programme *(cont.)*

The success of the Solar Night Schools Programme in preventing these children from being excluded from school depends to a large extent on a modality of integrated service delivery that is premised and built on a network of mutually supportive initiatives in the community served by the Barefoot College. Despite very limited resources and a hugely challenging environment, the programmes are sustained by their integration with and mutual support of each other, which produces a virtuous cycle compounding its own success. The Solar Night Schools are integrated within a network of other development programmes that support the children and their families; the programme is decentralised so that it can better respond to different contexts – a structure which further embeds it in the community; and an intercultural education modality makes education relevant not only in terms of the mainstream curriculum but also for the children and their families. Besides receiving education, students benefit from health services, communication resources, toys and learning materials provided by other development programmes operated by the organisation. These include the production of the solar-powered lanterns that enable the operation of the schools at night; the building of rainwater harvesting tanks adjacent to each night school, which helps to ensure that families in semi-desert areas will still be provided with water by their children – even if they attend school during the hours they would otherwise spend collecting it from wells; and the provision of community education about, for example, the advantages of sending girls to school, through puppet theatre (historically an integral part of the community's cultural traditions) and radio programmes broadcast from the Barefoot College. This integrated model also allows for the sharing of funds among different projects to support one another, and for the shared provision of materials, personnel, training and infrastructure across the different projects. Some examples of this are the fact that many night schools' alumni are incorporated into the Barefoot College's development projects. Alumni work as, for example, solar engineers, coordinators of craft workshops or of the local early childhood care and education centres, cooks for the Barefoot College community, or as cultural workers in the community.

Further examples of this integration among projects lie in the training offered to teachers by the Barefoot College's Health Centre to identify common health issues in their students, and in the vocational training that the children receive. The Barefoot College's Health Centre provides care for the children and their families. Dental care is also provided. Technicians trained by the Barefoot College check local water supplies for dangerously high levels of fluoride in the water. Families contribute substantially, further embedding the programmes in the community. Local communities generally provide the buildings for the night schools and other activities (such as fairs, workshops and meetings), and contribute voluntarily with cash, food, time or work to the realisation of the programmes. Supervision and management of the Solar Night Schools is largely by Village Education Committees and a Children's Parliament, both volunteer organisations run by the community. The degree of ownership that the communities have of the programme further enhances its sustainability, given that development interventions are generally successful to the extent that they are appropriated by and integrated into the communities where they are targeted. This embedding of the Solar Night Schools in a mutually supportive and integrated network of initiatives grounded in the Barefoot College is what has precipitated and sustained the success of the programme. The complexity generated by multiple integrated and mutually supportive initiatives is probably what has increased exponentially the chances of success of these programmes over the potential of, say, one isolated initiative to get poor, rural, low caste girls from illiterate families into school.

Sustainable change requires interventions at all levels inside and outside the education system

It is worth noting that complexity theory is in many perspectives akin to dynamical systems theory: one aspect of the shift in social theory from structuralism to complexity theory involves this focus on the dynamic, on the constantly evolving, where

structuralism has been criticised for its rigid, static and ahistorical perspectives on phenomena. Another key aspect of the shift from structuralism to complexity theory involves a shift from the reductionist perspectives typical of the former – as in, say, "Angela's failure is because, as a black girl from a low socio-economic status background, she is a member of the underclass" – to the focus on dynamically emergent phenomena typical of the latter. In the perspective of complexity theory, Angela's failure is not pre-determined, and it is not reducible to what neo-Marxist structuralists might have called her class, race and gender. Rather, a myriad factors and dynamics are involved, some compounding each other, others mitigating each other, in a multi-dimensional, iterative and recursive conception of causality. Trying to isolate the influence of a particular factor either in explaining failure or in effecting change is not only impossible, but also wrongheaded. Perhaps the major practical conclusion to draw here is that changing education systems to rid them of their inequities and inefficiencies requires massive interventions at all levels of the system, to which it should be added that substantial interventions at many levels of the society in which the education system is embedded would also be a prerequisite for sustainable change.

My focus here has been on the perennial question of what is effective in the sustainable change of a failing or under-performing education system. This would require intervention, as I have argued, at every possible level, including factors associated with the state and its education and economic policies, and possibly factors beyond even the grasp of the state – those that are associated with the forces and consequences of globalisation, for example. They would include factors associated with school leaders and teachers, with the students themselves, with their parents, with the curriculum, with schools' organisation, with the local community – the list is, if not endless, long indeed. But, given that I indicated earlier that complexity theory enables little or no causal relationship to be predicted from a knowledge of initial conditions to emergent phenomena, how can we know what to do about each of these factors? If it is both impossible and wrongheaded to try to isolate and assess the importance of any one factor, how can we even know in which direction we should try to push any factor? Fortunately, what we know from research in education gives us quite a few clues. The fact that complexity theory has less predictive utility than we might wish does not negate education's research findings. This is because we are talking about two different spheres with very different levels of complexity. We know, for example, that feedback provided to learners on the appropriateness of their constructions of new knowledge has an immensely powerful effect on learning. We can predict with substantial confidence that learners who receive feedback on the soundness or otherwise of the inferences they have drawn in the process of learning will learn more effectively than those who do not. Complexity theory's relative lack of predictive utility doesn't undermine our confidence in predicting this outcome, because this is not in itself a particularly complex phenomenon. Complexity theory does not apply here (at least not at the level at which we are discussing the phenomenon). Complexity theory has to do with complex systems, and it is at this level that it lacks strong predictive capacity.

We know that parental involvement in their children's learning enhances learning; that good school leaders create effective learning environments through good management practices; that poor children provided with a school lunch learn more effectively than students who do not benefit from such a policy; that students who are likely to find employment learn more effectively than those who perceive little likelihood of work. If we know all this, and can predict with a reasonable degree of confidence an improvement in learning outcomes in each of these domains, then surely we can predict

that change in the direction of enhanced learning outcomes in each of these domains is more likely to aggregate, in a complex adaptive system constituted by all of these factors (and more), to enhanced learning than to decreases in learning outcomes across the school?

Complexity theory thus indicates, in other words, that what it might take to change a system's inertial momentum from an ethos of failure to one of sustained development is massive and sustained intervention at every possible level, until the desired change emerges from this new set of interactions among these new factors and sustains itself autocatalytically. And despite complexity theory's relative inability to predict the direction or nature of change, we are, by implementing at each constituent level changes whose outcome we can predict with reasonable confidence, at least influencing change in the appropriate direction and thus stand a better chance of effecting the desired changes across the complex system as a whole.

Massive and sustained intervention at every possible level demands, unfortunately, very substantial resources. If there are many failing schools in a country's education system, choices might have to be made about where resources should be targeted. Trying to spread whatever resources are available across all failing schools may well result in the effects of the investment simply being dissipated. In each school, in other words, the intervention will have been too meagre to make any impact on the prevailing inertial momentum. Each school will in all likelihood revert to its ethos of failure, with the resources wasted. It may therefore be necessary to target the available resources at only a few selected schools for maximum impact – which is what it will probably take in terms of the arguments I have presented here. This will of course increase the level of inequity in the education system, a consequence that is morally questionable. But as yet I can see almost no way around this.

One might in response to this conundrum select the target schools based on a criterion that may reduce levels of inequity: for example, one might select, say, the thousand worst performing schools, or those schools that are attended predominantly by students from the poorest homes. Or one could select schools that are attended predominantly by students from minority groups (if those minority groups are indeed the least well off or in other ways excluded). The additional challenge in these cases is, of course, that these schools are going to be the hardest to turn around, and will demand substantial additional resources. The question then arises as to how policy makers might be able to predict which schools are more likely to change under the impact of massive and sustained intervention and investment of resources.

Complexity theory and educational research for sustainable change

Murray Gell-Mann, Nobel laureate in physics, has offered the caution that complexity focuses necessarily on "coarse-grained" (1994, pp. 29-30) descriptions and explanations of systems whose self-organising intra- and inter-actions normally render them too complex to be encapsulated by the standard repertoire of (educational) research tools, unless the complexity of the phenomena is abstracted and reduced to a workable level of statistical generalization. Paul Cilliers (2005) has noted that the sheer scope of the variables within complex systems makes modelling them a tricky, if not impossible, task. Such models would have to be as complex as the original, since the distributed, non-linear features of complex systems do not easily allow for the compression of data. My own view, as I have indicated, is that complexity is best used as a metaphor in which to understand the nature of systemic continuity and change.

As I have argued, in this perspective there are no independent interventions: proposed changes at the classroom level, for example, have implications at school and district levels (for example, for teacher development, parental expectations, school resources, accountability, and so on) and need to be supported by related interventions across multiple levels. Most important is a change in the paradigms of our thinking about research on education: away from input-output "black-box" causal models to modelling the specific, local linkages that actually interconnect actors, practices, and events across multiple levels of organisation; and away from single interventions and simplistic solutions to the recognition of the need for coordinated changes throughout the system and to its constraining and enabling contexts and resources.

To conclude by way of a restatement of what I see as the most important insight of complexity theory with regard to sustainable change and development in education: it is that new properties and behaviours emerge not only from the elements that constitute a system, but from the diversity and myriad connections among those elements. The successive linear addition of new elements multiplies exponentially the number of connections among the constituent elements. It is in this shift from linear to exponential orders of magnitude, but of course only in systems of incredible scale, that the power of complexity theory lies. The concepts of emergent phenomena from a critical mass, associated with notions of lock-in, path dependence, and inertial momentum, contribute to a perspective on continuity and change that indicates what conditions might need to be in place for the emergence of sustainable, positive, system-wide change and development in education.

Notes

1. The consideration of complexity theory offered here and this paper itself are based on fuller explications provided in Mason (2008a), "Complexity theory and the philosophy of education", and in Mason (2008b), "What is complexity theory and what are its implications for educational change?". Interested readers who seek to engage further with the field might consider these two articles, and also the chapters in Mason (ed.) (2008c), *Complexity Theory and the Philosophy of Education*.
2. The concept of inertia, most commonly used in physics, is probably familiar to social scientists less in association with the concept of momentum and more in terms of its association with resistance to movement, viz., the "inertial mass" of a heavy object on a high-friction, level surface. The concept of inertial momentum is, however, also common in physics, denoting, in rather over-simplified terms, the resistance of an object in motion to changes in its velocity. See Mason (2008b) for an introduction of the term to the social sciences.

References

Anderson, P.W. (1972), "More is different", *Science,* Vol. 177/4047, pp. 393-396.

Arthur, W.B. (1989), "Competing technologies, increasing returns, and lock-in by historical events", *The Economic Journal,* Vol. 99/394, pp. 116-131.

Arthur, W.B. (1990), "Positive feedbacks in the economy", *Scientific America,* Vol. 262, No. 1, pp. 92-99.

Cilliers, P. (2005), "Knowing complex systems", in K. Richardson (ed.), *Managing Organisational Complexity: philosophy, theory, and application*, Information Age Publishers, Greenwich, CT, pp. 7-20.

Lorenz, E.N. (1963), "Deterministic nonperiodic flow", *Journal of Atmospheric Science,* Vol. 20/2, pp. 130-141, http://dx.doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

Gell-Mann, M. (1994), *The Quark and the Jaguar: Adventures in the Simple and the Complex*, W. H. Freeman, New York.

Harvey, D. (1990), *The Condition of Postmodernity*, Blackwell Publishing, Oxford.

Kauffman, S.A. (1992), *Origins of Order: Self-organisation and Selection in Evolution,* Oxford University Press, Oxford.

Mason, M. (2008a), "Complexity theory and the philosophy of education", *Educational Philosophy and Theory*, Vol. 40/1, pp. 4-18.

Mason, M. (2008b), "What is complexity theory and what are its implications for educational change?", *Educational Philosophy and Theory,* Vol. 40/1, pp. 35-49.

Mason, M. (ed.) (2008c), *Complexity Theory and the Philosophy of Education,* Wiley-Blackwell, Oxford.

Prigogine, I. (1980), *From Being to Becoming,* W. H. Freeman, San Francisco.

Van Twist, M. et al. (2013), "Coping with very weak primary schools: Towards smart interventions in Dutch education policy", *OECD Education Working Papers*, No. 98, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k3txnpnhld7-en.

Waldrop, M. (1993), *Complexity: The Emerging Science at the Edge of Order and Chaos,* Viking, London.

Chapter 3.

Hierarchies, networks and improvisation in education governance

Henno Theisens

The Hague University for Applied Sciences

Over the past three decades, major trends have transformed the context of educational governance and created new governance challenges. Partly in response to these trends new forms of governance have risen, relying less on strong centralised rational planning and more on decentralised actors and market forces. These new forms of governance have not always solved existing problems and sometimes created new problems. However, because of societal changes, returning to a strong central government with rational planning is no longer possible.

This then raises the question what the next governance innovation should be, moving beyond the state and the market. While some propose governance networks as a promising avenue, the horizontal nature of networks creates tensions with the vertical, hierarchical organisation of ministries. This makes the position of civil servants working on the intersection of these vertical and horizontal logics of networks and hierarchies and their ability to cope with the tensions between them very important.

Introduction

In the past three decades the position of central governments in OECD countries has changed significantly. Power has moved away from central governments in different directions: upwards, towards international organisations, sideways to private institutions and non-governmental organisations and downwards towards local governments and public enterprises such as schools and hospitals. Where once we had central *government*, we now have *governance,* which can be defined as the processes of establishing priorities, formulating and implementing policies and being accountable in complex networks with many different actors (Theisens, 2013). This is a general trend in all OECD countries that ranges across different policy systems. It is certainly the case in education, but also health, public safety, welfare and other fields are all to a greater or lesser extent touched by it.

These changes are a direct result of policy decisions throughout the 1980s and 1990s, decisions that were part of a new way of thinking about the role of government: New Public Management (NPM). At its core, NPM is about the belief that governments are not just the solution to all kinds of social problems, but often also part of the problem. The assumption is that taxation, regulation and public production reduce the power of markets and enterprising individuals to come up with solutions and innovations for social problems. NPM is often used to describe a cluster of policy initiatives that are all in some way or another aimed at reducing the size and impact of government and increasing the operation of markets (Politt and Bouckeart, 2011).

NPM has been hotly debated and often rightly so. Privatisation and introducing markets has not always had positive effects and sometimes has had negative effects (Waslander et al., 2011). But while this critique is correct, it misses the point that NPM is also a reaction to a very real challenge, that modern societies have become too complex to be governed by strong central governments through rational public planning. Manuel Castell puts it very well: “The nation state has become too big for the management of everyday life and too small to control global flows of capital, trade, production, and information” (Castells, 1998).

In this chapter I will argue that, while NPM is by no means perfect, a return to a strong central government and rational planning is not feasible either. I will argue that instead of focussing on structures and incentives, which both traditional government approaches and NPM are prone to do, it is more important and fruitful to focus on the types of individuals, particularly their competences and skills, which are populating these governance structures.

Shifting societies: More global, more liquid and more interdependent

Strong central governments have been important factors in the growth of welfare and wellbeing throughout the 19th and 20th century. They ensured the rule of law and provided stability. They build railroads, highways, schools and hospitals and ensured public access to these public services. Strong central governments build the welfare state, with support for the sick, the unemployed and the elderly.

So why are these formerly extremely successful strong central governments becoming less and less effective? Strong central governments are not inherently bad, but the societies of which they are a part have changed profoundly. And in this new societal context strong central government no longer functions as effectively as it used to. These

societal changes can be framed in many different ways, but here the focus is on three clusters of changes: our societies are becoming more global, more liquid and more interdependent. Each of these trends affects the effectiveness of central government in different ways (Theisens, 2013).

More global – global financial markets, international organisations and social media

The world has become much more integrated in the past 30 years. Globalisation – the deepening, widening and speeding up of global interconnectedness – has meant that it is more and more difficult to consider national states as closed systems (Held and McGrew, 2007). The most extreme example of this are global financial markets which are already operating like one, global real-time system. But the markets for goods and services too are increasingly global, hampered less and less by national borders and facilitated by low cost of global transportation and communication. Global communication of course is greatly facilitated by the Internet, which provides an enormous capacity for global data exchange at very low costs.

Globalisation has significant consequences for the governance system of national states. One of the most important impacts is the decreasing influence of political power, especially the power of national politics, which is deeply imbedded in the nation state. The essence of national political power is still territorial, legitimated by democratic elections of a territorially bound electorate. The essence of modern power, most prominently that of large investors, is the fact that they are not bound to any territory. They can move their investments across the globe almost without obstacles. If these investors do not like the national tax regime or find the quality of the workforce wanting they can easily shift their stock portfolios or even their investments. To a lesser extent this is true for production companies, which have invested in production facilities that cannot be relocated without incurring high cost. But even for these companies every new major investment means an opportunity to relocate across national boundaries. This flexibility across national borders limits the power of elected politicians: as national wealth and employment critically depend on the presence of these kinds of companies (Bauman, 2000). Another important impact is the fact that economic competition is now to a large extent global, implying that countries need to worry about how internationally competitive their national economies are. This worrying translates into a political agenda that limits taxation, stimulates labour market flexibility and education of world class quality. The political agenda of national governments in other words is limited by the forces of globalisation.

But globalisation is not merely an economic phenomenon; it is a social and political phenomenon too. Partly in response to *economic* globalisation a process of *political* globalisation has developed, with a growing number of influential international organisations. Typical for these organisations in a globalised world is that they are no longer merely platforms where countries are making international agreements but that they are more and more – often without democratic legitimation – directly influencing national policies and the activities of actors within countries. This further limits the authority of national governments.

Political globalisation is not just about international organisations; it is also about the power of individual citizens to organise themselves horizontally and across borders through social media and the Internet. Social media and the Internet help to inform individuals, organise movements and publish the activities and ideas of these movements

to the world at large. While the Arab spring was not caused or even successful due to Facebook and Twitter, it is undeniable that they played an important role. And while the Occupy movement may not have achieved all that it set out to accomplish, the organisation of this global movement was definitely helped by these social media (Gladwell, 2010).

More liquid – decreasing influence of traditional organisations and changing social structures

Globalisation is a highly visible large-scale process but *inside* OECD countries there are profound changes too. Perhaps the most important of these is the decreasing influence of the traditional institutions, rules and practices that governed human life and the increasing stress on individuals to shape their own lives. In the past three decades, strong national institutions – governments, political parties, unions and churches – have lost much of their power and their leaders have lost a lot of their self-evident authority (Giddens, 2000). At the local level changes are obvious as well. The traditional village, with its strong sense of community, its rules and social control is disappearing (Mak, 1996). Even closer to home, traditional nuclear families, long considered the corner stone of societies are changing as well; the model of the nuclear family is joined with myriad other models: more people for example are living alone, more married people are living without children, more unmarried people are living with children and there are more single parent families (Carnoy, 2001; OECD, 2013).

The fabric of society has changed profoundly and the essence of this change is that things are less solid and more fluid (Bauman, 2000). Individual choice has increased tremendously and this has made the job of governing extremely complex. In an attempt to deal with this complexity, governments across the OECD have decentralised authority towards organisations like local governments, hospitals and schools. These organisations oversee only a small part of the system, reducing the complexity they need to take into account. They are also closer to individual citizens, making it possible – at least in theory – that they can take the preferences of these citizens into account.

More interdependent – traditional structures, horizontal networks, and modern technologies

The combined forces of globalisation and increasing individual freedom might project the image of an open space where individuals are freely and individually moving around. However, the world is not just more global and more fluid, it is increasingly interdependent too.

Traditional institutions and communities have been replaced by more flexible and horizontal networks where individuals are often a member of different, overlapping professional and social networks. In these networks, individuals cooperate, share information and relax; often these functions are mingled too. These networks are strongly facilitated by information and communication technologies (ICTs), through online communities and platforms for co-operation. Social and digital developments are mutually reinforcing each other.

For governments this has consequences. Governments need to govern a society in which fleeting, horizontal networks are now an important phenomenon. Individuals are more independent vis-à-vis traditional institutions, but they are quite capable of taking collective initiative through horizontal networks. These initiatives that are not formed

through regular channels are difficult for governments to understand and to respond to. Moreover, these networks make it relatively easy for citizens to organise opposition against government plans. For example, one well-formulated tweet can lead to serious public debate.

Government reform and reformed government

At the same time that societies have changed, governments have changed too. Since the early 1980s – the era of Reagan and Thatcher – governments across the OECD have attempted reforms of varying degrees of depth, width and success (Politt and Bouckeart, 2011; Laegreid and Christensen, 2011). The general underlying rationale for these reforms was the notion that the big governments that came into being with the rise of the welfare state since the early 1960s were no longer just solutions to societal problems, but the cause of social problems as well. Increasingly, governments were perceived not only as inefficient and ineffective, but, moreover, as slow, wasteful and as a barrier for entrepreneurial individuals and innovation. Reagan once quipped: "the 10 most dangerous words in the English language are, 'Hi, I'm from the Government, and I'm here to help'" (Reagan, 1988).

New Public Management – increasing efficiency and coping with complexities

These reform programmes are often inspired by the loose bundle of concepts and programmes labelled as New Public Management (NPM). NPM set out to reduce the size and influence of governments and replace this with – theoretically – more efficient markets or market type mechanisms. Flowing from this general principle are diverse measures such as: lowering taxes, privatising public services, decentralising the authority of public services, granting increased autonomy to public service providers, increasing competition between public service providers, introducing performance indicators and using these to steer the public system.

From the beginning, the ideas of New Public Management have been debated (Dijkstra, 2012). Criticisms include the special responsibility of a democratically legitimated government for such things as equality, equity and other public values, which cannot be left to the market. Other criticisms are about the new inefficiencies flowing from the use of performance indicators, which only focus on measurable types of performance. It has also been debated whether or not decentralised public services that operate in competition are really more efficient. For example, schools now often compete for students, but does this make them better schools or better advertisers? Are they innovating or merely copying each other?

Much of the criticism of New Public Management is warranted. However, what should not be forgotten is that these reforms were not just self-standing attempts of governments to become more efficient, but also a reaction to societies that had changed profoundly. Interpreting these reforms in that light offers another picture.

Privatisation and decentralisation are not just about raising efficiency. They can be interpreted as ways in which national governments are moving power to places better suited to handle the complexities of global, liquid and interdependent societies: they unburden national governments and leave organisations in charge that only have to focus on one specific service (e.g. privatising telecom companies) a smaller territory and fewer citizens (decentralising to local governments) or both (as in the case of giving more autonomy to schools). So the critique of opponents of New Public Management that

decentralisation and fragmentation have led to a governance system that is fragmented is not wrong, but it is incomplete. For the reasons mentioned above, the alternative to a fragmented governance system is not a centralised system: complex societies cannot be ruled rationally from one centre, if only because the amount of information that needs to be processed to make that possible far outstrips what any central government can achieve.

Likewise, increasing competition between services is not just about efficiency but is also a way to increase the diversity of the services on offer and allowing increasingly demanding citizens to choose between these different options. In almost all OECD countries the freedom for parents to choose their child's school has increased. And while there has been debate about the negative effects of this development: larger inequities and costs as certain schools are catering to too few children; there is no serious discussion of abolishing parental choice. Parental choice is by and large perceived as of *intrinsic* value.

Performance indicators, finally, are perhaps even more widely debated. Opponents claim that assessing schools and teachers based on indicators and making the results public, carries serious risks such as teaching to the test and schools focussing on a limited number of competencies that are measured by these tests (see Smith, Chapter 4). They also argue that school performance indicators suggest that schools can be easily compared while in fact the socio-economic background of children explains a lot of the variation. *However*, in societies where the traditional authority of schools, school principals and teachers has eroded, some form of performance measurement to reassure parents of the quality of education their child receives is necessary. It is no longer enough that teachers say children are doing well, or that the head of school asserts that education standards are excellent. Performance indicators fill the gap that is left when the authority of teachers and school leaders decrease and parents still want to know how their child is doing in school.

Horizontal networks and hierarchical government

On the one hand, the critique of New Public Management is legitimate. Many of these reforms have not brought the kind of efficiencies that their proponents claimed and they have had (in some cases serious) side effects. On the other hand, there is no simple way back. Traditional centralised governance structures simply won't work in contemporary societies. The big question is: what's next? Are there new forms of governance, new structures, new processes that allow governments to move beyond NPM *and* traditional centralised government?

Since the early nineties, the term "network" has been presented as a promising concept (Thompson et al., 1991; Hufen and Ringeling, 1990). The idea is that governments should embrace the fact that there are many different actors involved in governance and co-operate with these actors to govern. There are some good arguments for this idea. Networks operate on the basis of links between different actors and are in tune with the growing interdependence of society. Networks are more flexible than the traditional hierarchical organization of the state and therefore fit the dynamics of "liquid modernity". Networks operate on the basis of trust. They function because people are willing to co-operate and sacrifice short-term gains for the benefit of long-term co-operation. They are very different in this sense from markets and hierarchies, where the need for trust is minimised through complex systems of incentives and rules (Cerna, 2014). This is not just a nice conceptual thought: the Nobel laureate Elinor Ostrom has shown through decades of empirical research that in the absence of strong central control and powerful market forces, local networks under the right conditions can effectively solve shared problems, like maintaining complex irrigation works (Ostrom, 2010).

A crucial question for governments is how to relate to these networks. Roughly, there are two roles. The first role is that of a government hovering above the actors in a network. This involves creating the arena within which networks of public and private parties operate: establishing frameworks, formulating a strategic vision, facilitating knowledge and feedback and operating as a crowbar when participants in a network arrive at a stalemate (Pierre and Peters, 2005). A second role is when government – often through its officials – itself acts as a player in networks. At the edges of the government, where ministries, civil society organisations, private companies and citizens come together there are dynamic networks that address social problems. These networks are most visible at the local level, where networks of citizens, local NGOs, private companies and local governments are tackling issues around safety, health and education. But the national government too is a player in international, national and local networks around social themes (Pierre and Peters, 2005; Steen et al., 2010).

One of the problems of the network concept is that we understand relatively little about how interactions in networks lead to systems that are sufficiently organised to function. Proponents of the idea point to swarms of starlings, or schools of fish or ant colonies that without centralised control perform very complex acts of co-ordination. But these are simple systems, with animals that by and large operate instinctively. These systems have been successfully modelled using computers and providing large amounts of model birds or fish with a few simple individual behavioural rules. Both societies and human beings are more complex. It is simply not clear how order is created in complex networks, notwithstanding the heavy borrowing of some public administration scholars from the work on chaos and complexity theory (Boutelier, 2011; see also Mason (Chapter 2) for an example of how this might apply in education).

The governance challenge: combining the vertical and the horizontal

The reality of contemporary governance is one of two organizational principles that are difficult to combine. First, a vertical line, from the minister downwards runs a ladder where all involved in education can be placed on a rung. The logic of this vertical line, a hierarchy, is well known and has been well studied from the classical work by Weber in late 19th century onwards (Roth and Wittich, 2013). Hierarchy is a convenient and rational way to organize people and to organise decision-making. It allows for both specialisation and the integration of specialised activities by different experts. There is a reason why so many of our organisations, from ministries, to schools and hospitals to corporations are structured in hierarchies. Within the government there is yet another reason for this: the bureaucratic organisation of a ministry ensures that – at least symbolically – all public officials ultimately fall under the minister – a minister who in turn is accountable to a democratically elected parliament.

At the same time there is now the logic of the horizontal line: horizontal networks between all kinds of different stakeholders in the education system, the ministry, the inspectorate, teacher unions, all kinds of interest organisations, advisory groups and schools themselves. These horizontal networks lack the clarity of the vertical, hierarchical line but they are indispensable in the fragmented governance systems that decentralisation, privatisation and deregulation have led to. The setting of priorities, the formulation and implementation of policies require many parties to work together. While this is important, there are often no clear structures for this type of horizontal co-operation.

This means that policy makers and professionals at all levels of the education system are both part of a hierarchy and of one (or many) horizontal networks. Moreover, both these vertical and horizontal organisational principles are important at the same time: civil servants for example need to work together with others horizontally, while getting approval for actions vertically from those higher up in the hierarchy. The need to be flexible, to negotiate and to show initiative for horizontal networks to work is often at odds with the need for prior approval, limited freedom and a tendency to punish failure that hierarchies are prone to. This tension is felt at the level of managers too. Typical hierarchical performance evaluation takes place at the individual level. How do you make the performance of civil servants in horizontal networks visible, how do you manage and reward them on the basis of this information? The tension between these vertical and horizontal logics is the context within which policy makers at all levels must operate.

Policy makers and civil servants in a context of hierarchies and networks

The fact that policy makers must now operate in a context where vertical and horizontal principles are playing out at the same time is reflected in many contemporary analyses of the competencies that civil servants require. Where traditionally civil servants were expected to be loyal, neutral, efficient, bureaucratic, honest and knowledgeable, modern civil servants must also be conscious of the (political) context, result oriented, flexible and collaborative (Hart, 2014; Steijn, 2009; Niessen, 2001).

These analyses are interesting but they are too narrowly defined. First, policies are no longer made by civil servants alone but are created in networks where many actors interact. We should be concerned with the competencies of all policy makers involved. Second, merely adding a number of competencies may not be enough. Working on the intersection between vertical and horizontal modes of organisations means that policy makers (including civil servants) are increasingly required to play a new role; a role that is sometimes referred to as "boundary spanners" (Williams, 2010; Steen et al., 2010). Boundary spanners work together with other professionals outside the silos of their hierarchical organisations to ensure that social problems are addressed. This requires them to work within the hierarchical logic of their own organisation and within horizontal networks *at the same time*. This implies at least two things:

- Given the aforementioned tensions between hierarchies and networks it requires policy makers to make constant judgements about what type of behaviour is effective and appropriate in which settings and how to balance these two logics at any given time.
- If governance takes to a large extent place in horizontal networks, then a much deeper understanding of how these networks operate and how policy makers can contribute to them is essential.

The remainder of this chapter will focus on those two issues. This is not a field where there is an abundance of information, let alone evidence available. In the Public Management research group at The Hague University for Applied Sciences we have begun to explore a number of concepts what we believe may be helpful. What follows is not so much a recipe book for making good policy, but a number of insights and research questions that we are currently working on. It is in other words, work in progress.

Practical wisdom to make judgements in complex situations

An ethnographic study at the Dutch ministry of the interior confirms the idea that civil servants need to make constant judgements. A researcher was given permission to observe the inner workings of the Ministry of the Interior and the behaviour of civil servants. First findings, that have not been published yet, show that in almost all meetings he attended, questions about the role of the ministry were on the table. And that this role was very much addressed both in terms of vertical logic: what does the minister/ our director-general want, and horizontal logic: what do we have to offer to the local professionals who are primarily responsible solving this issue?

In balancing these two logics there is not "one right answer". Instead this requires making judgements in complex situations with different, sometimes opposing, values that are all equally important. Where can civil servants be innovative and where should they respect the primacy of democratically elected politicians? Where should they act in their role as representative from the minister, top down ensuring that all citizens receive equal treatment, and where should they be more flexible, supporting local governments and schools to do what is best in their particular context? Civil servants need a moral and political compass to navigate these difficult waters.

What civil servants need is the old Aristotelian notion of "practical wisdom": the ability to make ethical decisions in complex settings and to translate these decisions in effective actions (Schwartz and Sharpe, 2010). This is not about classical integrity of civil servants, which is very much about abiding to the rules and conforming to norms. This is about coming up with effective actions in situations where values like loyalty, transparency and effectiveness may be at odds with one another. What is true of civil servants is true for policy makers more in general; policy makers whether inside or outside the ministry need some practical wisdom.

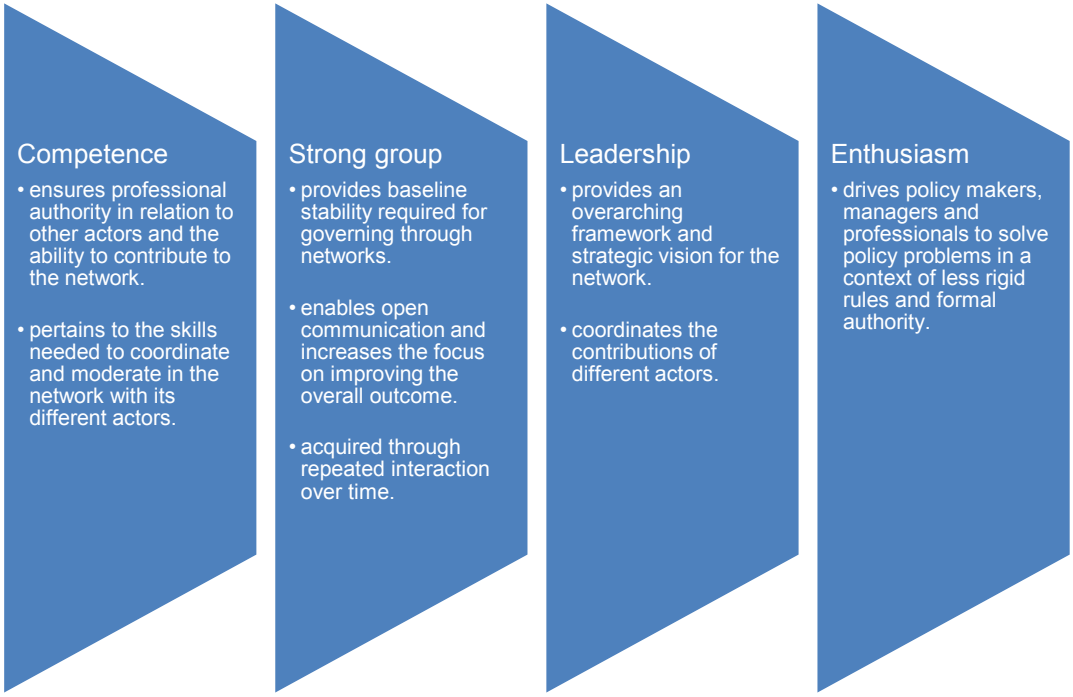
Practical wisdom is a professional competence that professionals learn in years of training and experience and within the context of a professional learning community. One of the problems is that our ministries and many of the other organisations involved in policy making are not exactly professional learning communities. In fact many of the traits of these organisations possess are actually hampering the development and use of practical wisdom. Social psychology shows that whenever rules or protocols are introduced there is a tendency for people to stop worrying about the matter at hand and start focussing on the rules (Schwartz and Sharpe, 2010). The same is true for the incentive systems that New Public Management has introduced at all levels of government. People worry more about the incentives than the actual issues (Schwartz and Sharpe, 2010).

Improvisation – what governance can learn from jazz musicians

The second conclusion is that a much deeper understanding of how networks operate and how civil servants can contribute to them is addressed through a different line of research. One of the more interesting books written on the phenomenon of governance in contemporary societies focuses on the role of improvisation (Boutelier, 2011). The basic tenet of the book is that social order in borderless, fluid societies is no longer a given, but that regardless of this OECD societies are not descending into chaos. Order is constantly being made and remade by the actions of individuals; much like jazz musicians who create seemingly effortlessly meaningful and often beautiful music on the spot by improvising. Our research group takes this idea of improvisation one step further and focuses on the simple question: what can civil servants learn from jazz musicians?

Quite a bit is known about the art of improvisation in jazz music (Berliner, 1994). One of the more important insights is that while the act of improvisation happens in the spur of the moment, there is a long list of conditions that must be met for improvisation to be possible (see for an overview Hartog, 2014). The following key elements can be distilled from the literature (Figure 3.1):

Figure 3.1. Factors required for effective improvisation



Source: Author's own work, based on Vogelaar (2014), Marsalis (2008), Barrett (1998), Berliner (1994).

Competence: Improvisation requires the individual players to be technically proficient with their instrument. They must have mastered the jazz tradition, through learning from and replaying of other musicians, and they need to understand musical theory and practice such as harmonics, melody, rhythm and chords (Berliner, 1994). And they need to be self-reflective, knowing about their strengths, weaknesses and what they can bring to improvisation (Barrett, 1998). Similarly, effective civil servants in horizontal networks need to be professionals who know what they are talking about, they cannot simply coordinate other people's work. They must be knowledgeable to be taken seriously by the others in the network and to be able to make contributions. At the same time, they need to have a particular set of skills, their instrument, which they take to the network.

Strong group: Improvisation requires rehearsal. Rehearsal helps the members of the group to know each other, understand each other and learn to listen to each other. Practice builds routines, more or less predictable responses that help structure the on-stage improvisation. On a deeper level, improvisation presupposes a cohesive group of musicians willing to communicate openly (both sharing and listening) and able to sacrifice personal ideas and self-expression to improve the quality of the overall outcome (Marsalis, 2008; Berliner, 1994). This means that any governing through networks that does not have some level of stability is doomed to fail. The actors in horizontal networks need to get to know each other; this takes time.

Leadership: While improvisation seems to imply that nobody is in charge, this is rarely the case. Most often one of the musicians puts together the band, selecting the individual musicians, creating a concept and a particular sound. But during improvisation this is true as well: one of the musicians is usually in charge of deciding what the framework for the improvisation and who gets a chance to play solo, etc. (Berliner, 1994). Though the term "horizontal" suggests that all are equal, this most often cannot be the case if outputs are to be expected. There needs to be someone in a leadership role and given the position of the government, this most often will be a civil servant. The type of leadership is very different though. The jazz metaphor suggests that leadership may be as much about preparation: framing the issues at hand, carefully choosing the participants in the network. During network interactions leadership ensures that there is an overarching framework for the process and skilfully coordinates the contributions of different actors.

Enthusiasm: difficult to grasp but very important is the notion of enthusiasm. Enthusiasm within every musician as he plays, enthusiasm generated between musicians as they perform on stage and enthusiasm in the audience as they listen to and become part of a performance, in turn inspiring the musicians (Vogelaar, 2014). The ability to generate enthusiasm is not particularly well understood, but in horizontal network and more importantly in an age where rules and authority are crumbling, enthusiasm may be what drives policy makers, managers and professionals to solve pressing policy problems.

The notion of improvisation is used here as a metaphor, with all the strengths and weaknesses that it entails. Our research group will be looking to explore further this metaphor along two lines. Practically we want to bring together civil servants and jazz musicians in workshops to see whether and how they can learn from each other. Scientifically we are interested in the research questions that this metaphor can generate. These are research questions that can then help to establish the thinking on much stronger empirical foundations.

Conclusion

This chapter has analysed developments in the societal context of governance and governance itself over the past three decades. It concludes that profound changes have taken place, new forms of governance, i.e. NPM, have arisen, and that these new forms have not always solved existing problems and sometimes have created new problems. However, because of societal changes, returning to a strong central government based on rational planning is no longer possible. This raises a question regarding what the next governance innovation should be, moving beyond the state and the market.

The growing importance of networks and their horizontal nature creates tensions with the vertical, hierarchical organisation of ministries and other organisations involved in educational policymaking. Neither of these two organisational logics is likely to go away anytime soon. This makes the position of policy makers working at the intersection of these vertical and horizontal logics of networks and hierarchies very interesting. How do these actors deal with these tensions?

When thinking about complex education systems we tend to focus on structures, rules, indicators and evidence. But these mean nothing if the people at all levels inside these systems cannot use their professional judgement, their practical wisdom to decide which rules are important and when exceptions must be made, which indicators are worthwhile in which contexts and how to weigh evidence.

Moreover, in complex education systems, as in other complex systems, working together across the boundaries of organisations is both a necessity and a challenge. It is necessary because without the input of civil servants, labour unions, inspectorates, school managers and teachers policies will not adequately reflect the multifaceted nature of the system. It is a challenge because all these actors represent different perspectives and are embedded in different organisations each with their own logic and interest. Working together across organisational boundaries to make and implement policies requires the skills of an improvisational artist: like a strong sense and knowledge of tradition, the ability to listen and share, and the enthusiasm to pull people together.

In other words, to govern complex education systems we need to educate those who govern to do so effectively in complex contexts. Teaching them to use their practical wisdom and improvisation may be a small step towards that.

References

Bauman, Z. (2000), *Liquid Modernity*, Polity Press, Cambridge.

Barrett, F.J. (1998), "Managing and improvising: lessons from jazz", *Career Development International*, Vol. 3/7, pp. 283-286.

Berliner, P.F. (1994), *Thinking in Jazz: The Art of Improvisation,* University of Chicago Press, Chicago and London.

Boutellier, H. (2011), *De Improvisatiemaatschappij, Over De Sociale Ordening Van Een Onbegrensde Wereld* [The Improvisation Society, Creating Social Order in a Borderless world], Boom Lemma Uitgevers, The Hague.

Carnoy, M. (2001), "Work, society, family and learning for the future", in *What Schools for the Future*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264195004-en.

Castells, M. (1998), "The education of city planners in the information age", *Berkeley Planning Journal*, Vol. 12/1, pp. 25-31.

Cerna, L. (2014), "Trust: What it is and why it matters for governance and education", *OECD Education Working Papers*, No. 108, OECD Publishing, Paris, http://dx.doi.org/10.1787/5jxswcg0t6wl-en.

Dijkstra, G. (2012), *Botsende Waarden in en rond Publieke Organisaties*, Intreerede [Conflicting Values in and around Public Organisations, inaugural lecture], The Hague University of Applied Sciences, The Hague.

Giddens, A. (2000), *Runaway World*, Routledge, London.

Gladwell, M. (2010), "Small change, why the revolution will not be tweeted", *New Yorker Magazine*, www.newyorker.com/magazine/2010/10/04/small-change-3, (accessed 23 March 2015).

Hart, P. (2014), *Ambtelijk Vakmanschap 3.0: Zoektocht Naar Het Handwerk Van De Overheidsmanager* [The Craftmanship of Civil Servants 3.0: In Search of the Skills of the Public Manager], Netherlands School of Public Administration, The Hague.

Hartog, M. (2014), "Towards improvisational governance? Jazz improvisation and networked complex governance", paper presented at the 7th Euro Mediterranean Dialogue on Public Management (MED7), 8-10 October 2014, Rome.

Held, D. and A. McGrew (2007), *Globalisation Theory, Approaches and Controversies*, Polity Press, Cambridge.

Hufen, J. and A. Ringeling (1990), *Beleidsnetwerken, Overheids, Semi-Overheids – En Particuliere Organisaties In Wisselwerking* [Policy Networks, Public, Semi Public and Private Organisations in Interaction], VUGA Uitgiverij, The Hague.

Kwakkelstein, T., A. van Dam and A. van Ravenzwaaiij (eds.) (2012), *Van Verzorgingsstaat Naar Waarborgstaat, Nieuwe Kansen Voor Overheid En Samenleving* [From Big Government to Big Society, New Chances for Government and Society], Boom Lemma Uitgevers, The Hague.

Laegreid P. and T. Christensen (2011), *Transcending New Public Management: The Transformation of Public Sector Reforms*, Ashgate Publishing Ltd, Farnham, United Kingdom.

Mak, G. (1996), *Hoe God Verdween uit Jorwerd, Een Nederlands Dorp in de Twintigste Eeuw* [How God disappeared from Jorwerd, A Dutch Village in the Twentieth Century], Uitgiverij Atlas, Amsterdam and Antwerp.

Marsalis, W. (2008), "Why teach jazz?", in R. Miles and R. Carter (eds.), *Teaching Music Through Performance in Jazz,* GIA Publications, Chicago.

Niessen, C.R. (2001), *Vluchten Kan Niet meer, Iets Over Beleids – En Beheersproblemen En Hun Oplossing* [Fleeing is Futile: About Policy and Control Problems and their Solution], Vossiuspers UvA, Amsterdam.

OECD (2013), *Trends Shaping Education 2013*, OECD Publishing, Paris, http://dx.doi.org/10.1787/trends_edu-2013-en.

Ostrom, E. (2010), "Beyond markets and states: polycentric governance of complex economic systems", *American Economic Review*, American Economic Association, Vol. 100/3, pp. 641-672.

Pierre, J. and B.G. Peters (2005), *Governing Complex Societies: Trajectories and Scenarios*, Palgrave Macmillan, London.

Pollitt, C. and G. Bouckaert (2011), *Public Management Reform: A Comparative Analysis New Public Management, Governance, and the Neo–Weberian State,* 3rd ed., Oxford University Press, Oxford.

Reagan, R (1988), "Remarks to Representatives of the Future Farmers of America", Speech, July 28, 1988, www.reagan.utexas.edu/archives/speeches/1988/072888c.htm (accessed 23 March 2015).

Roth, G. and C. Wittich (2013), *Max Weber, Economy and Society*, University of California Press, Oakland.

Schwartz, B. and K. Sharpe (2010), *Practical Wisdom*, Riverhead, New York.

Steijn, B. (2009), "Over de competenties van de nieuwe' ambtenaar" [About the competencies of the "new" civil servant], in: *Rijksambtenaren Van de Toekomst,* [Civil Servants of the Future], Ministry of the Interior and Kingdom Relations, The Hague, pp. 30-45.

Theisens, H. (2012), *Governance in Interessante Tijden, Een Essay* [Governance in Interesting Times, an Essay], The Hague University of Applied Sciences.

Thompson, G. et al. (1991), *Markets, Hierarchies and Networks, The Coordination of Social Life*, Sage Publishers, Thousand Oaks.

Van der Steen, M., R. Peeters and M. van Twist (2010), *De Boom en Het Rizoom, Overheidssturing in een Netwerksamenleving* [The Tree and the Rhizome, Government Steering in a Network Society], Netherlands School of Public Administration, The Hague.

Vogelaar, R. (2014), *De enthousiasme Trilogie, Flame, Flow, Flood* [The Enthusiasm Trilogy: Flame, Flow and Flood], Adfo groep, Amsterdam.

Waslander, S., C. Pater and M. van der Weide (2010), "Markets in education: An analytical review of empirical research on market mechanisms in education", *OECD Education Working Papers*, No. 52, OECD Publishing, Paris, http://dx.doi.org/10.1787/5km4pskmkr27-en.

Williams, P. (2010), "Special Agents: The nature and role of boundary spanners", Paper to the ESRC seminar series: Collaborative futures: New insights from the intra and inter-sectoral collaborations, University of Birmingham.

PART 2.

ACCOUNTABILITY

Chapter 4.

Exploring accountability: National testing policies and student achievement

William C. Smith

RESULTS Educational Fund

The recognition that all children have the right to a quality education has pushed education provision and quality assurance to the top of policy agendas. The use of test scores for accountability purposes has risen worldwide, accompanied by a belief in the market model (e.g. school choice) as a strong way to ensure and monitor quality education. There is an open question however about how effective these market forces are, and whether the use of test scores is achieving the desired improvements in education performance. This chapter uses the National Testing Policies (NTP) outlined by Smith (forthcoming) to explore common practices found in schools in educator based testing for accountability systems, providing policy-makers with a rich illustration of school practices in each NTP.

Introduction

The use of large-scale standardised tests is engrained in education systems around the world. As the presence of testing[1] in our schools appears to be widely accepted, how test scores are used and disseminated continues to be hotly contested. As education systems globally adopt policies that hold educators accountable for their students' test scores, research is being conducted to inform policymakers of the potential benefits and consequences of such a system. Underlying the movement toward more educator-based accountability is a belief that accountability pressure will shape educator behaviour, creating a more efficient and effective system. Student test scores, aggregated at the classroom or school level, are then used as an accepted measure of educational quality.

This chapter explores the mechanisms underlying the relationship between educator-based accountability and student achievement, as measured by test scores. To identify countries which practice educator-based accountability, National Testing Policies (NTP) are used. National Testing Policies were recently introduced by Smith (forthcoming) and placed participants of the 2009 data collection for the OECD Programme for International Student Assessment (PISA) on a rough continuum based on the presence and intensity of educator-based testing for accountability. After careful policy analysis Smith organised systems into three categories: Summative, Evaluative, and Formal Sanction/Reward. Using this established categorisation scheme, this chapter examines which school level practices are more likely to be found at increasingly intense levels of accountability and whether the positive relationship between educator-based accountability and student achievement, often reported in past research, are related to policies that may expand already identified equity concerns.

The chapter starts by describing the global expansion of testing and educator-based accountability. This is followed by a review of the past research on accountability and student achievement, which reveals on average a marginal, positive association between student achievement in systems that are under educator-based accountability pressure and those that are not. Caveats and limitations to previous research precede the section outlying the classification of accountability systems into National Testing Policy categories. Following a breakdown of the variables used in the analysis, the results of a three level hierarchical linear model (HLM) are provided graphically to address the research questions: (1) How does the incorporation of school practices and policies differ by National Testing Policy (NTP)? (2) Which of these school practices or policies are more responsible for the commonly found relationship between NTPs that promote educator-based testing for accountability and student achievement? The chapter ends with a brief conclusion that summarises the main findings and questions whether the benefits of educator-based accountability can be perceived to outweigh the exclusionary practices indicated by the analysis. The findings suggest that teacher-based accountability systems are capable of exerting the pressure to change teaching and school practices. However, the benefits of these changed practices for students are less clear. Presenting a worrisome development, the analysis finds a significant correlation between systems focussing strongly on testing for teacher accountability and practices to exclude students of lower socio-economic background and low achieving students.

Global expansion of educator-based testing for accountability

The second half of the 20th century saw an increasing number of countries use large-scale standardised tests in their education system. This led Phelps to conclude at the turn of the century that there is a "clear trend towards adding, not dropping testing programs" (Phelps, 2000: 19). This trend expanded beyond the borders of the industrialised countries as educational reformers looking for the "magic bullet" to strengthening education insisted that "improving national (or state) testing systems is an important, perhaps the key, strategy for improving educational quality" (Chapman and Snyder, 2000: 457). The speed at which countries adopted national test policies increased early in the 21st century. In their examination of educational systems between 1995 and 2006, Benavot and Tanner (2007) found that countries incorporating annual national tests into their education program more than doubled from 28 to 67.

Although testing is widely practiced, countries use tests and test results in different ways. Some countries use testing to gauge the national health of the education system, while others focus attention on using test scores to evaluate schools, teachers or students. Additionally, a few nations (e.g. Panama and Greece) have decided against testing, or did not have a national testing policy as of 2009 (Smith, 2014b). The ideal of education for all and recognition that all children have the right to a quality education can be seen to promote the use of test scores for accountability purposes (Mundy, 2006; Rose, 2005). Test scores are seen as objective comparable information that is essential for parents to act as customers in an educational market (Smith, 2014a). Over the past 20 years "the development and implementation of accountability systems has been one of the most powerful, perhaps the most powerful, trend in education policy" (Volante, 2007: 4).

The belief that every child has the right to an excellent education and that excellence or quality can be measured objectively through standardised tests has prompted a movement toward educator-based testing for accountability. Testing for accountability involves the "application of formal or informal, positive or negative consequences on educators dependent on their students' performance measures" (Smith, 2014a: 6). Global and country peer pressure legitimating testing for accountability as an acceptable use of student test scores together with the belief that market competition will lead to greater efficiency and thereby increased quality (Chubb and Moe, 1990) prompts countries to turn toward more intense applications of accountability. Given the commonly placed faith in science to measure education quality through student test scores (Smith, 2014a), the greater efficiency gains promised by a competitive market should be apparent in student achievement scores.

Accountability and student achievement

Three formal meta-analyses have been conducted to estimate the effect of accountability systems on student test scores. All identify positive associations (i.e. greater accountability associated with higher student test scores) ranging from a marginal (Belfield and Levin, 2002) to medium effect size (Phelps, 2012). They are:

- Belfeld and Levin (2002), in a meta-analysis of 25 studies that examined the link between competitive pressure and educational outcomes in the United States found a modest effect on scores with a 1 standard deviation increase in between school competition associated with a 0.1 standard deviation increase in test score.

- In an examination of 14 studies exploring the effects of test-driven external accountability systems, Lee (2008) found small positive effects when mathematics and reading scores were averaged but no effect on the racial achievement gap. Lee concluded that the marginal mean effect size for school accountability "does not lend strong support for claims for school accountability" (p. 616).
- Phelps (2012), in his investigation of the effects of testing on student achievement over a 100-year period found the largest effect size. However, the 160 studies included in his study were not limited to those addressing accountability issues and he notes that the largest effect sizes were associated with tests that provide timely feedback, a trait more commonly associated with in-class formative assessments. Finally, in a qualitative literature review of studies that apply an incentive structure that emphasises explicit consequences (e.g. the No Child Left Behind Act in the United States), Elliot and Hout (2011) find positive significant effects are concentrated in studies that examine elementary grade mathematics and that the mean effect size (0.08) was substantially lower than that needed to close the achievement gap between the United States and its peer countries.

While most of the research on testing for accountability has focused on the experience of the United States, the global trend toward similar accountability practices suggest other systems can learn from their experience. Past research finds the effect of testing for accountability differs by ethnic group, student ability, subject and type of accountability (Figlio and Loeb, 2011). Some studies find testing for accountability disproportionately disadvantages ethnic minorities (Hanusheck and Raymond, 2005), while others suggest it closes the Hispanic-white achievement gap while increasing the black-white gap (Hanushek and Raymond, 2004), or that it has unilateral benefits for minorities (Carnoy and Loeb, 2003). Looking across three international datasets, Woessman (2004) concluded that the effect of external exams on student achievement also indicates a relative advantage for higher ability students. Differences in effect are also present across subjects as research using a multitude of comparison strategies finds a positive effect of testing for accountability on mathematics test scores with a weaker or non-significant effect found on reading scores (Cronin et al., 2005; Dee and Jacob, 2009; Figlio and Loeb, 2011; Lee, 2008; Wong, Cook and Steiner, 2009). A similar trend is found when persistence in achievement gains are explored (Chiang, 2009). Cronin et al. (2005) suggest that differences between subjects may be due to the dependency of mathematics understanding on classroom instruction, while reading is relatively more easily affected by parental involvement.

The type of testing for accountability applied can also lead to divergent results. Studies suggest that policies that publish comparable school level results have a positive effect on student test scores, although the "practical significance of this gain is negligible" (Springer, 2008: 5). Additionally, in systems that use explicit consequences to deter poor performing schools, student achievement is found to be higher (Dee and Jacob, 2009). When these two approaches to testing for accountability are compared, the benefit of using explicit consequences is greater than the benefit of applying market pressure through publishing school level results (Bishop et al., 2001; Hanushek and Raymond, 2005). In comparing eighth grade student test scores for students in the United States before and after explicit consequences were added to the accountability system, Hanushek and Raymond (2005) find an approximately 0.1 standard deviation increase in scores on the National Assessment of Educational Progress. As the publication of results

was not a significant factor in test scores, the authors concluded that the increase in scores can be attributed to the explicit consequences. The gains in test results in testing for accountability system may be partially due to increases in higher expectations of all student groups, including students with disabilities, in the tested subjects (Ysseldyke, Dennison and Nelson, 2004).

Caveats and limitations to previous research

As illustrated in the diversity of effects above important caveats must be taken into account before drawing conclusions about the relationship between educator-based testing for accountability and student achievement. First, the presence of a positive effect is often reliant on the level of local autonomy the school has. This has led some research to conclude that educators must be free to respond to the demands of parents in order for an accountability system to function effectively (Woessman, 2007). Second, the positive relationship between educator-based testing for accountability and student achievement appears to be dependent on other contextual and demographic factors including the subject tested and the grade and ethnicity of the student. Finally, and perhaps most importantly, how the preponderance of past studies define, and therefore classify, educator-based accountability is problematic.

Previous studies on school accountability are often limited by how they distinguish which schools are subject to accountability pressures. A large number of studies use a single school level practice (whether or not a school publicly posts their test results) as the parameter for inclusion (see Rosenkvist, 2010, for review of studies). The use of this single school level practice to define school or educator-based accountability has at least two unmet assumptions. First, asking school principals (who are generally the educational actors asked in national and international assessments to capture school climate) assumes that the public dissemination of school aggregate test scores are the responsibility of the school. Instead, this task is often completed by the national ministry of education or regional educational authority. While some principals and school leaders may be taking this into account when answering the question, it is clear not all are aware of the publication of results or believe it is their position to acknowledge that when asked about school climate. For example, in a pooled sample of 22 countries that require school level test results to be made available to the public, less than 50% of principals acknowledge that these aggregate results are publicly posted (Smith, forthcoming).

The second problematic assumption is the belief that the environment that will produce accountability pressure can be captured by using information from a single school. The publication of school level results by a single school does not provide enough information to verify whether parents, in their role as customers in the education market, are provided with adequate comparable information. This comparable school level information is necessary for parents to create an environment where schools feel accountability pressure and therefore adjust their practice as needed. Instead, to capture which schools are subject to an environment that mandates educator-based testing for accountability national policies must be examined.

The importance of looking at national testing policies

An emphasis on national testing policies is important given the rapid spread of within-country standardised testing and the ability of national policy to capture the overarching accountability environment. National testing policies leverage educator

behaviour by creating an environment conducive to some behaviour and unfavourable to others. Decisions on testing and how testing should be used remains highly centralised where "national ministries of education typically act as agents imposing this activity [testing] on schools and education systems" (Kamens and McNeely, 2010: 6). Even in highly decentralised systems with historically high levels of school autonomy, such as the United States, national policies have been shown to influence classroom and school decisions and streamline educator practices (Booher-Jennings, 2005; Luna and Turner, 2001). Potentially homogenous educator practices within a single country leaves some to suggest that commonly reported single country studies (see Figlio and Loeb, 2011) provide limited insight into the mechanisms of accountability (Woessmann, 2007).

To distinguish between national policies and provide a rich description of what educator-based testing for accountability looks like at the school level, recent cross national research has proposed the use of National Testing Policy categories (Smith, forthcoming, 2014a). National Testing Policies (NTPs) classify countries based on how they use standardised tests to hold educators accountable for their students test scores (Table 4.1).

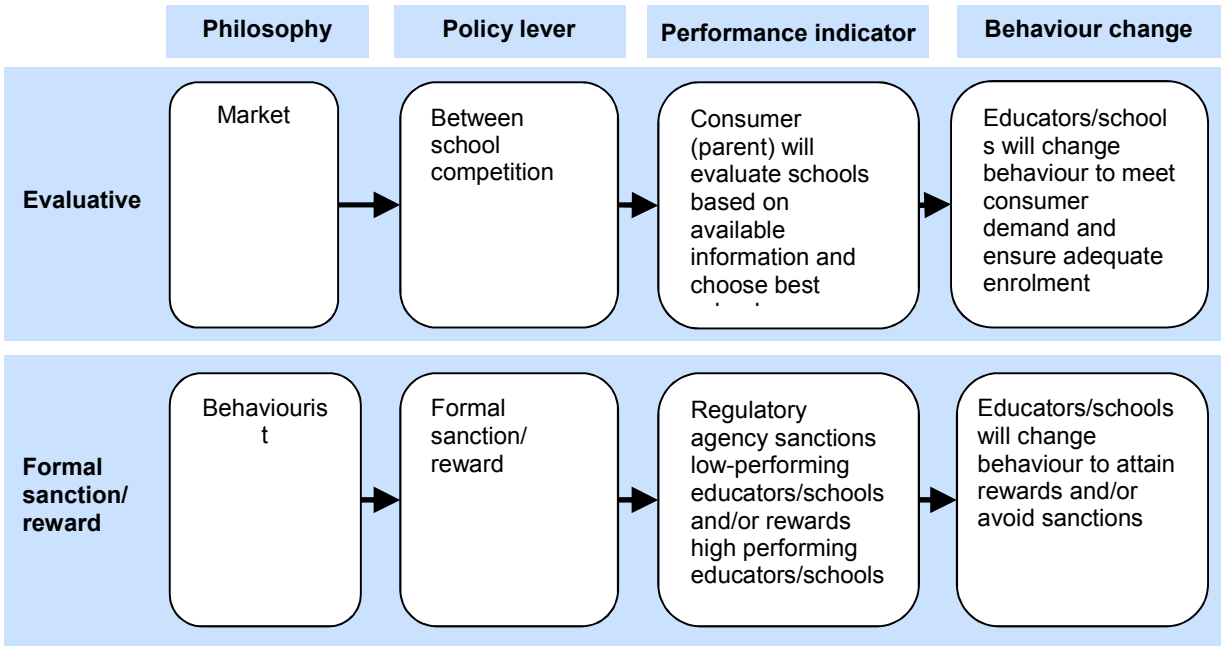
Table 4.1. National testing categories

National Testing Policy	Summative	Evaluative	Formal Sanctions and/or Rewards
Description of National Testing Policy	The use of national or regional examinations as a tool that summarises student learning and is shared with parents; when disseminated is done so at the national or regional level.	The use of national or regional examinations as a tool that summarises student learning and is disseminated at the school level to allow for between school comparisons.	The use of national or regional examinations as a tool that summarises student learning, is disseminated at the school level, with school/class level results used to apply rewards or sanctions.

Source: Smith (forthcoming).

An important distinction between Summative countries and Evaluative or Formal Sanction/Reward countries is the level of test score aggregation. By aggregating student level test scores at the national or regional level, Summative countries do not match schools or educators directly with their student test scores and therefore do not practice educator-based testing for accountability. In contrast, both Evaluative and Formal Sanction/Reward systems hold educators accountable, although the motivation behind each differs. As illustrated in Figure 4.1, Evaluative systems are based on a market philosophy designed to create competition through the production of comparable information. The end goal is an adjustment in educator practices to align with the demands of consumers and ensure their livelihood by maintaining adequate enrolment. In contrast, Formal Sanction/Reward systems, based on a behaviourist philosophy, add a level of intensity to the accountability pressures already present in Evaluative systems. By making educator performance akin to student test scores and linking it to direct sanctions and/or rewards, Formal Sanction/Reward countries expect educators to alter their practices to avoid punishment or seek rewards.

Figure 4.1. Behaviour change model for national testing policy categories



Source: adapted from Smith (forthcoming).

Empirical analysis

The analysis in this chapter is drawn from Smith (forthcoming) which classifies the participants of the 2009 PISA test into the three NTP categories outlined in Figure 4.1. Categorisation was completed through an in-depth analysis of national and international policy documents and follow up validity checks by national experts. For more details on NTP categorisation as well as the technical specifications and supporting documents from which this chapter is drawn, see Smith (forthcoming). Using the pooled sample of countries into NTP categories allows for cross-policy comparisons, informing policy makers as they decide how to incorporate tests and accountability into their educational system, and capturing the heterogeneous student outcomes that are associated with different approaches to testing (Harris and Herrington, 2006).

With National Testing Policies established, this chapter explores educator-based testing for accountability by examining two questions.

1. How does the incorporation of school practices and policies differ by NTP?
2. Which of these school practices or policies are more responsible for the commonly found relationship between NTPs that promote educator-based testing for accountability and student achievement?

School Practices and Policies

To address these questions, emphasis is placed on school level variables extracted from the supporting questionnaires of the 2009 PISA test. In addition to school type (private = 1, public = 0) specific school policies and practices used in this analysis include:

- School Monitoring: School uses student assessments to monitor the schools progress from year to year (1=yes, 0=no).

- School Comparison: School uses student assessments to compare the school with other schools (1=yes, 0=no).
- Parent-School Comparison: School provides aggregated and comparable school results to parents (1=yes, 0=no).
- Publicly Posted: School achievement data is publicly posted (1=yes, 0=no).
- Principal Evaluation: School achievement data is used to evaluate the principal's performance (1=yes, 0=no).
- Teacher Evaluation: Student achievement data is used to evaluate the teacher's performance (1=yes, 0=no).
- Admission Decision: School uses student achievement, including test scores in their admission criteria (1=sometimes or always, 0=never).
- Transfer Decision: School is likely to transfer out student due to low academic achievement (1=likely or very likely, 0=not likely).
- Standardised Tests: School completes two or more standardised tests per year (1=yes, 0=no).
- Time in Math: Continuous measure of how many minutes per week the student reports spending in mathematics classes. Taken from the 2009 PISA student questionnaire, aggregated to the school level using student survey weights, and grand mean centred.
- Extra-curricular Activities: Includes two standardised latent variables identified through principal component analysis using Varimax rotation (see Appendix H in Smith (forthcoming) for factor scores):
  - Academic Extra-curricular Activities: Standardised factor identifying the availability of academically focused extra-curricular activities (i.e. Math Club).
  - Non-academic Extra-curricular Activities: Standardised factor identifying the availability of non-academically focused extra-curricular activities (i.e. Sports Team).

Student level variables

Mathematics score in PISA 2009 was chosen as the student achievement variable in this analysis as past research indicates a greater association between educator-based testing for accountability and mathematics, relative to other subjects. The use of students' math achievement, therefore, provides a conservative lower-bound estimate, suggesting that if differences are not found between NTP and math achievement they are unlikely to be found between NTP and other subjects.

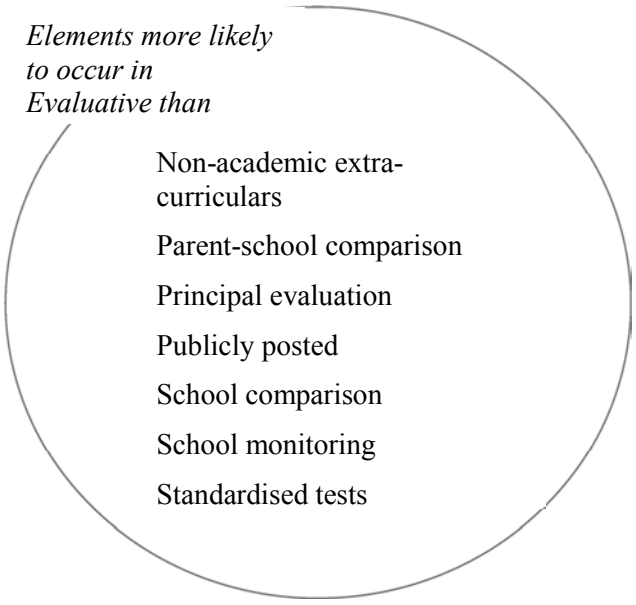
To take into account demographic and family background differences, four student level control variables are included. These include student gender (1=female, 0=male) and socio-economic status (SES) (taken from PISA's index of economic, cultural, and social status), as well as immigrant status (1=first or second generation immigrant, 0=native) and home language (1=primary home language is not test language, 0=primary home language is test language).

Describing schools in Evaluative and Formal Sanction/Reward systems

To address research question one – how does the incorporation of school practices and policies differ by NTP? – bivariate analyses were conducted to identify whether the school policies and practices outlined above are more likely to present in specific NTP systems. This basic analysis provides an interesting description of what the average school within the identified system looks like. Comparisons for both Evaluative and Formal Sanction/Reward systems are made to Summative systems (i.e. Summative is the reference group) through one-way ANOVA or chi-square tests, as appropriate, and only statistically significant differences ($p<.05$) are reported.

Figure 4.2 illustrates the school practices or policies that are more likely to be present in Evaluative systems relative to Summative systems. Compared to schools in systems that do not use school aggregate test scores, schools in Evaluative systems are more likely to participate in all six school accountability practices (school monitoring, school comparison, parent-school comparison, publicly posted, principal evaluation, teacher evaluation). This is not unexpected and it reinforces the belief that national policy can shape school level behaviour. Additionally, schools in Evaluative systems are more likely to take two or more standardised tests annually and offer more non-academically focused extra-curricular activities. The later result is somewhat surprising as past research often finds a narrowing of resources toward testing subjects and away from activities not associated with math, science and reading (Sterns and Glennie, 2010). Other research, however, has hypothesized that this increase is likely to occur as non-tested activities, such as band and theatre, are "crowded out" of the traditional school day and increasingly offered solely as an extra-curricular activity (Smith, forthcoming).

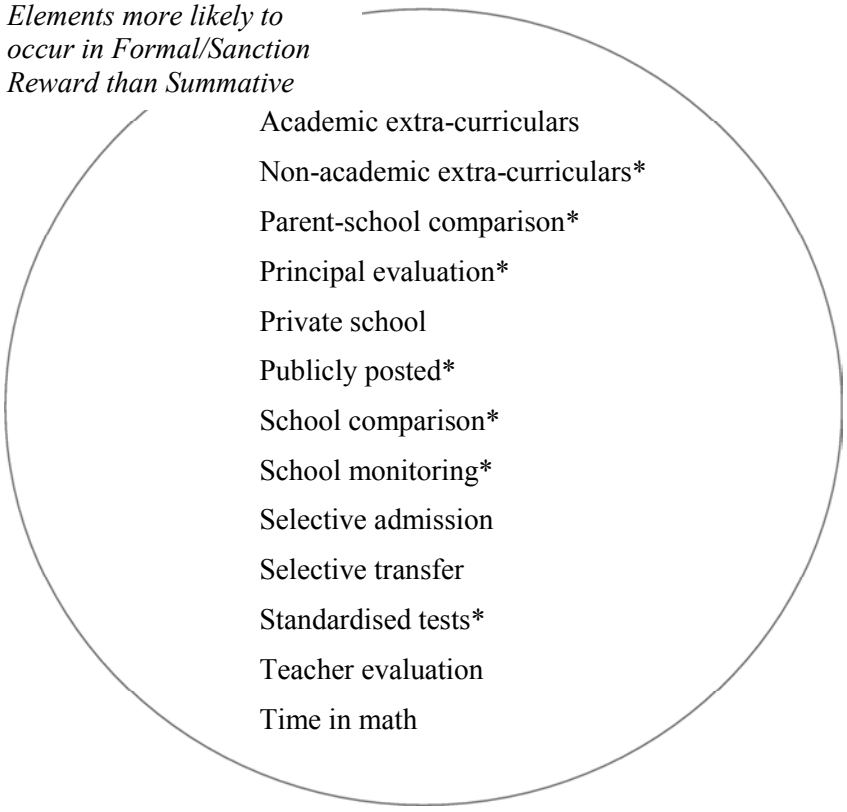
Figure 4.2. Likely procedures in schools using Evaluative rather than Summative systems



Source: Author's own work, based on Smith (forthcoming).

Figure 4.3 compares Formal Sanction/Reward systems and Summative systems using the same bivariate analysis. The figure illustrates that, relative to Summative systems, all school level practices included in this investigation are more likely to occur in Formal Sanction/Reward systems. In addition to the eight school practices more likely to occur in Evaluative systems, schools in Formal Sanction/Reward systems are: more likely to be a private school, spend more time on mathematics instruction, more likely to use student achievement as a criteria in their admission and transfer policies, and offer more academically focused extra-curricular activities. Among these significant comparisons the greatest differences are found in school comparison (found in 59.3% of school in Formal Sanction/Reward systems, 57.1%=Evaluative, 43.6%=Summative), teacher evaluation (64.7%=Formal Sanction/Reward, 61.6%=Evaluative, 47.1%=Summative), admission decision (66.6%=Formal Sanction/Reward, 55.8%=Summative), and standardised tests (35.3%=Formal Sanction/Reward, 32.2%=Evaluative, 26.8%=Summative).

Figure 4.3. Likely procedures in schools using Formal Sanctions/Reward rather than Summative systems



Note: Elements denoted with * are more likely in both Formal/Sanction Reward and Evaluative relative to Summative.

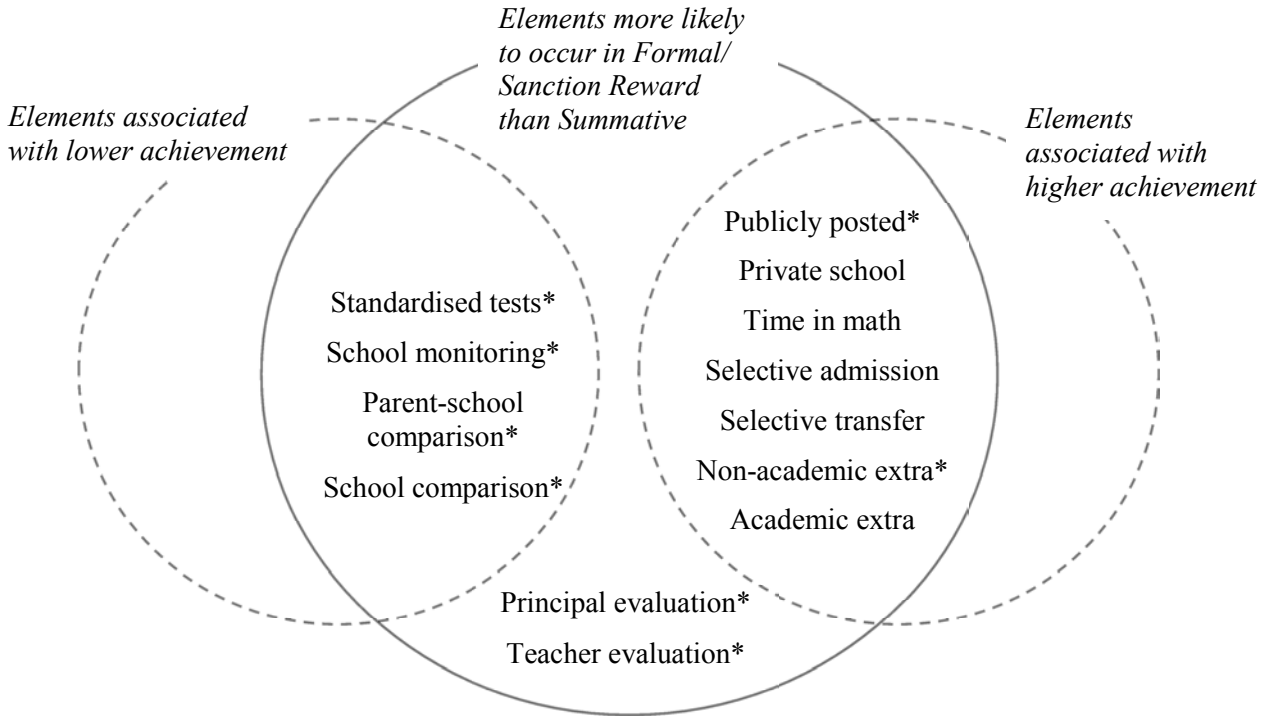
Source: Author's own work, based on Smith (forthcoming).

School practices associated with higher student achievement

Figure 4.4 uses the results of bivariate one-way ANOVAs examining the association between school practices and student achievement to create a Venn diagram overlapping those practices more commonly found in Evaluative and Formal Sanction/Reward

systems with those that are associated with student achievement. The figure identifies two important intersections: one with school practices that are more likely in Evaluative or Formal Sanction/Reward systems and are associated with *higher* student achievement, the other with school practices that are more likely in Evaluative or Formal Sanction/Reward systems and are associated with *lower* student achievement. Interestingly, using student achievement to evaluate principal or teacher performances are the only two practices more likely to occur in either system, and neither of these are significantly associated with student achievement. By breaking down the two intersecting circles, some understanding is gained of the likely differences in overall student achievement by NTP. The greater portion of school practices that are more likely in Evaluative systems (in denoted with * in Figure 4.4) and in the lower student achievement set suggests that student achievement in Evaluative systems may be lower than or not significantly different from student achievement in Summative systems. In contrast, the higher concentration in the higher student achievement set for Formal Sanction/Reward systems hints at a potential positive relationship between Formal Sanction/Reward systems and student achievement. Of these suggestive hypotheses, multi-level analysis predicting student achievement from NTP category confirms the initial positive relationship between Formal Sanction/Reward systems and student achievement. Student achievement scores in Evaluative systems were not significantly different from scores in Summative systems. As a result, the remainder of this analysis focuses on the relationship between school practices commonly found in Formal Sanction/Reward systems and associated with higher student achievement scores.

Figure 4.4. School practices associated with student achievement



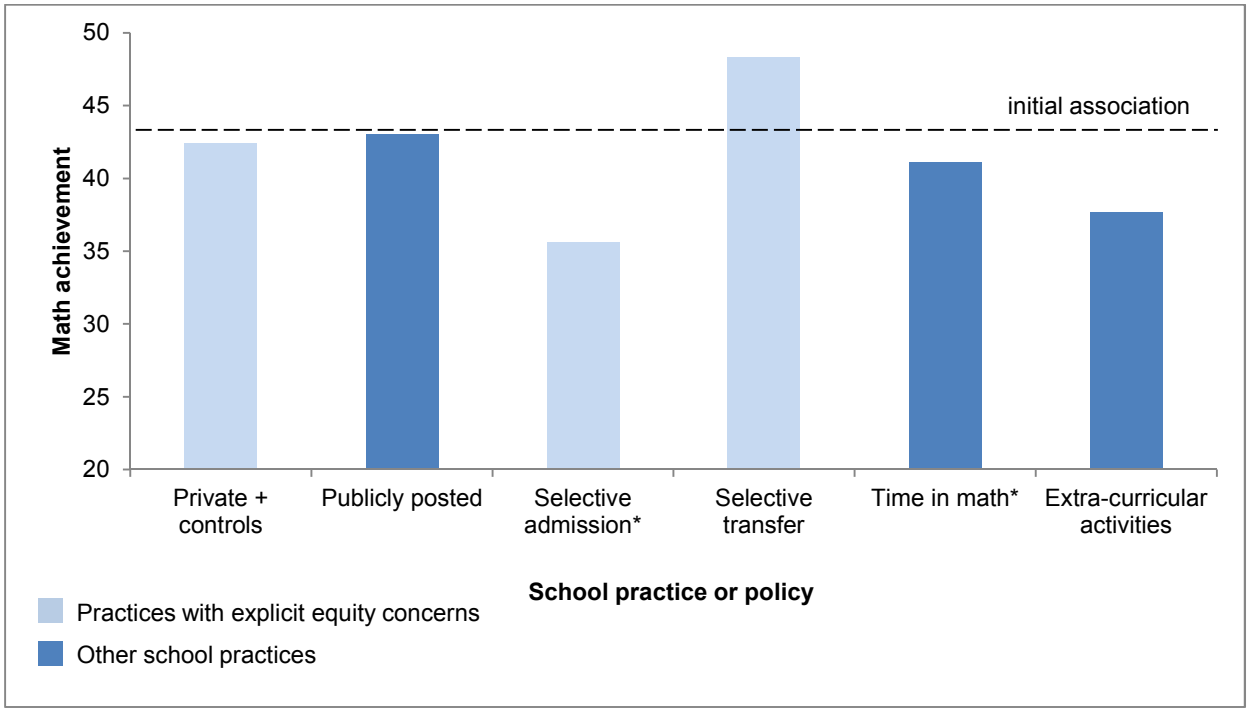
Note: Elements denoted with * are more likely in both Formal/Sanction Reward and Evaluative relative to Summative.

Source: Author's own work, based on Smith (forthcoming).

Can the increased likelihood of select school practices explain the relationship between Formal Sanction/Reward systems and student achievement?

To explore whether the seven school practices that are both more likely to be found in Formal Sanction/Reward systems and are related to higher student achievement can mediate the relationship between NTP and student achievement, results from a three-level random coefficient hierarchical linear model (HLM) are used (see Annex 4.A1: Regression results). The null relationship between NTP and student achievement indicates an initial positive relationship between Formal Sanction/Reward systems and higher math achievement with students in Formal Sanction/Reward systems scoring approximately
43 points higher than their peers in Summative systems do (dashed horizontal line). Figure 4.5 illustrates how the inclusion of different school practices mediates this initial relationship. The first bar in Figure 4.5 indicates that when school type (private vs. public) and student control variables are included in the model, the point difference decreases slightly to 42 points and remains significant. All remaining bars add alternatingly a single school practice to the model with school type and student controls. Non-academic and academically focused extra-curricular activities were included together in the last model.

Figure 4.5. Explaining the initial association between Formal Sanction/Reward systems and student achievement



Note: * Difference between Formal Sanction/Reward and Summative is not statistically significant ($p<.05$). Dashed line represents initial association between Formal Sanction/Reward and student achievement (43.574). Three level random coefficient HLM controlling for school type and student gender, SES, immigration status and home language. See Annex 4.A1 for respective regression results.

Source: Author's own work, based on Smith (forthcoming).

The results indicate that in public schools that do not selectively admit students based on their achievement there is not a statistically significant difference across NTP categories. Additionally, student achievement in public schools that spend the same amount of time on mathematics instruction is not significantly different from any other NTP category.

The school practices and policies included in Figure 4.5 can be divided into at least two categories: policies that directly shape the composition of the student body and those that dictate how schools engage with curriculum and test results. The first category, often identified in the literature as practices that shape the testing pool, brings with it equity concerns. In this study three (private, selective admission, selective transfer) of the seven levers for greater student achievement in Formal Sanction/Reward systems elevate school mean achievement by limiting their student body to those that can afford to attend or have the necessary achievement scores.

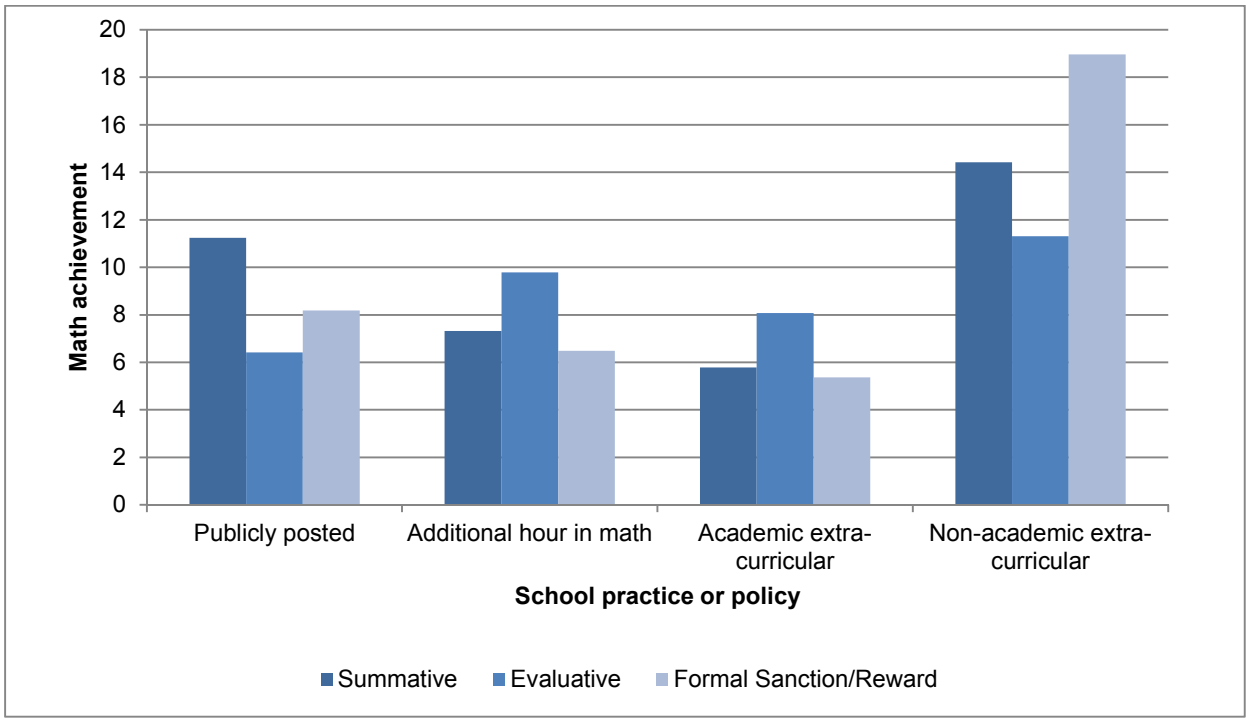
Past research on shaping the testing pool indicates that this is a fairly common practice in schools under accountability pressure. Shaping the testing pool generally excludes low performing students to increase the school's mean test score (Hanushek and Raymond, 2004). Similar to the results here, West, Pennell and Noden (1998) find that some schools participate in "cream skimming" by selecting in some students while selecting out others. Other research finds that schools shape the testing pool by increasing student retention (Hursh, 2005; Kornhaber, 2004a; Kornhaber, 2004b), transferring students into special education (Cullen and Reback, 2006; Jacob, 2005) or permanently excluding low achieving students (Figlio, 2006; Gillborn, 1996; Lewin and Medina, 2003). In New York City, schools altered the testing pool by categorizing low achieving students as transferred or working on their General Educational Development (GED) tests (Lewin and Medina, 2003). In the United Kingdom, Gillborn (1996) found that the permanent exclusion rate in schools increased 300% in the three years following the implementation of league tables. When government officials were asked to explain the dramatic increase, 8% attributed it to increased behavioural issues with 43% linked to increased between-school competition. Figlio (2006) investigated disciplinary records across multiple school districts in the four years surrounding implementation of Florida's accountability system. Focusing on the over 40 000 incidents where at least two students were suspended for the same event he found that harsher punishments were doled out to the lower performing student and that the difference in punishment increased during the testing season and among testing grades, a pattern not present prior to the accountability system. Figlio concluded that schools use discipline policies to reshape the testing pool by removing low performing students during testing periods through longer suspensions.

Balancing benefits with equity concerns?

Although the equity concerns involving school access are substantial, perhaps an increased return on achievement among the other four school practices – publicly posted test results, increased time in mathematics, greater availability of academically focused extra-curricular activities and greater availability of non-academically focused extra-curricular activities – would encourage policymakers to look past equity concerns, believing that the benefits outweigh the drawbacks. Figure 4.6 provides the different returns on student mathematics achievement associated with select school practices in each NTP category by inserting a cross-level interaction term into the three-level HLM. Results indicate that schools in Formal Sanction/Reward systems that participate in these school practices receive a marginally smaller return in three of the four practices examined. The non-significant interaction term indicates that there is not an increased

return on any of the practices for schools in Formal Sanction/Reward systems. In essence, this means that the association of student achievement with publicly posting school level results, increasing time in mathematics and increasing the availability of all types of extra-curricular activities is positive regardless of NTP category.

Figure 4.6. School practice and student achievement by testing (NTP) category



Note: Three level random coefficient HLM with cross level interaction controlling for school type and student gender, SES, immigration status and home language. Based on regression results of cross level interaction terms (see Annex 4.A1).

Source: Author's own work, based on Smith (forthcoming).

Conclusion

Both Evaluative and Formal Sanction/Reward systems appear to provide the accountability pressure necessary to change educator behaviour and school practices. The benefits of these alterations for students, however, are not entirely clear. For example, the greater amount of time spent on in mathematics instruction found in Formal Sanction/Reward systems is associated with higher student math scores. In terms of efficiency however, the analysis shows that the hours spent on mathematics instruction have similar impact on the PISA scores in mathematics across NTP categories. This indicates that while Formal Sanction/Reward systems may spend more time on math (with a positive effect on math scores), they are not necessarily more efficient with the time spent than the other NTP categories. As such, the argument put forth by proponents of education marketisation that sanctions and rewards are more efficient than other accountability mechanisms is not supported by the findings.

While student achievement is initially measured as higher in Formal Sanction/Review systems, controlling for policies that shape the testing pool renders the initial positive association with student achievement insignificant. This suggests that practices selecting

in and out students based on their achievement may be a major driver of differences in student achievement. Furthermore, past research found Formal Sanction/Reward systems to be more likely in displaying undesirable consequences such as higher dropout rates (Jacob, 2001; West and Pennell, 2000), higher teacher turnover in schools and a higher percentage of marginalised students (Clotfelter et al., 2004; Figlio and Loeb, 2011; Waterreaus, 2003), lending further weight to the suggestion that undesirable outcomes of coercive forms of educator accountability may outweigh potential benefits.

Note that this study cannot and does not make claims regarding the causality of the associations revealed. Those schools/systems that are facing difficulties (e.g. low student performance) might have introduced new policy/practice (e.g. formal sanctions and rewards) in order to improve their situation, while those schools/systems that are already doing well might not change their policy/practice. In order to address this, it would be important to consider the timing of policy implementation in addition to the variables already presented here. Further research on this topic could fruitfully explore this additional dimension.

In light of this study's findings, policy-makers should carefully evaluate the potential benefits against possible equity concerns when considering comparably coercive forms of accountability. The association of Formal Sanction/Reward systems with tendencies to exclude low SES students and low achieving students from comparative testing calls for a careful look at the consequences for broader goals of education, such as social inclusion and equal access to quality education.

Note

1. In this chapter, testing refers to large-scale standardised tests that are part of a school, district, regional or national testing programme.

References

Belfield, C.R., and H.M. Levin (2002), "The effects of competition between schools on educational outcomes: A review for the United States", *Review of Educational Research*, Vol. 72/22, pp. 279-341.

Benavot, A. and E. Tanner (2007), "The growth of national learning assessments in the world, 1995-2006", Background paper for the Education for All *Global Monitoring Report 2008: Education for All by 2015: Will We Make It.*

Bishop, J. et al. (2000), "The role of end-of-course exams and minimal competency exams in standards-based reforms", *Center for Advanced Human Resources Studies (CAHRS) Working Paper Series*, Vol. 4/1, Cornell University.

Booher-Jennings, J. (2005), "Below the bubble: 'Educational triage' and the Texas accountability system", *American Education Research Journal*, Vol. 42/2, pp. 231-268.

Carnoy, M. and S. Loeb (2003), "Does external accountability affect student outcomes? A cross-state analysis", *Educational Evaluation and Policy Analysis,* Vol. 24/4, pp. 305-331.

Chapman, D. and C. Snyder (2000), "Can high stakes national testing improve instruction: Re-examining conventional wisdom", *International Journal of Educational Development,* Vol. 20/6, pp. 457-474.

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

Chubb, J. and T. Moe (1990), *Politics, Markets, and America's Schools,* Brookings Institution Press, Washington, D.C.

Clotfelter, C. et al. (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,* Vol. 23/2, pp. 251-271.

Cronin, J. et al. (2005), *The Impact of the No Child Left Behind Act on Student Achievement and Growth,* Northwest Evaluation Association, Evanston, IL.

Cullen, J. and R. Reback (2006), "Tinkering towards accolades: School gaming under a performance accountability system", in T. Gronberg and D. Jansen (eds.), *Advances in Applied Microeconomics*, Emerald Group Publishing Limited, London, pp. 1-34.

Dee, T. and B. Jacob (2009), "The impact of No Child Left Behind on student achievement", *National Bureau of Economic Research Working Paper,* No. 15531.

Elliot, S.W. and M. Hout (eds.) (2011), *Incentives and Test-Based Accountability in Education,* National Academies Press, Washington, D.C.

Figlio, D. and S. Loeb (2011), "School accountability", *Handbook of the Economics of Education*, Vol. 3, Elsevier, Amsterdam.

Figlio, D. (2006), "Testing, crime, and punishment", *Journal of Public Economics*, Vol. 90/4-5, pp. 837-851.

Froese-Germaine, B. (2001), "Standardised testing + high-stakes decisions = educational inequity", *Interchange*, Vol. 32/2, pp. 111–130.

Gillborn, D. (1998), "Exclusions from school: An overview of the issues", in *N. Donovan (ed.), Second Chances: Exclusion from School and Equality of Opportunity,* New Policy Institute, London.

Hanushek, E. and M. Raymond (2004), "The effect of school accountability systems on the level and distribution of student achievement", *Journal of the European Economic Association*, Vol. 2/2-3, pp. 406-415.

Hanushek, E. and M. Raymond (2005), "Does school accountability lead to improved student performance?", *Journal of Policy Analysis and Management,* Vol. 24/2, pp. 297-327.

Harris, D. and C. Herrington (2006), "Accountability, standards, and the growing achievement gap: Lessons from the past half-century", *American Journal of Education,* Vol. 112/ 2, pp. 209-238.

Hursh, D. (2005), "The growth of high-stakes testing in the USA: Accountability, markets and the decline of educational equality", *British Educational Research Journal*, Vol. 31/5, pp. 605-622.

Jacob, B.A. (2005), "Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago public schools", *Journal of Public Economics*, Vol. 89/5-6, pp. 761-796.

Jacob, B.A. (2001), "Getting tough? The impact of high school graduation exams", *Educational Evaluation and Policy Analysis*, Vol. 23/2, pp. 99-121.

Kamens, D.H., and C.L. McNeely (2010), "Globalisation and the growth of international educational testing and national assessment", *Comparative Education Review*, Vol. 54/1, pp. 5-25.

Kornhaber, M.L. (2004a), "Assessment, standards, and equity", *Handbook of Research on Multicultural Education,* Vol. 2, pp. 91-109.

Kornhaber, M.L. (2004b), "Appropriate and inappropriate forms of testing, assessment, and accountability", *Educational Policy*, Vol. 18/1, pp. 45-70.

Lee, J. (2008), "Is test-driven external accountability effective? Synthesising the evidence from cross-state causal-comparative and correlational studies", *Review of Educational Research*, Vol. 78/ 33, pp. 608-644.

Lewin, T. and J. Medina (2003), "To cut failure schools shed students", *The New York Times*, 31 July 2003.

Luna, C. and C.L. Turner (2001), "The impact of the MCAS: Teachers talk about high-stakes testing", *English Journal*, Vol. 91/1, pp. 79–87.

Mundy, K. (2006), "Education for All and the new development compact", *International Review of Education,* Vol. 52/1, pp. 23-48.

Phelps, R.P. (2012), "The effect of testing on student achievement, 1910-2010", *International Journal of Testing,* Vol. 12/1, pp. 21-43.

Phelps, R.P. (2000), "Trends in large-scale testing outside the United States", *Educational Measurement: Issues and Practice*, Vol. 19/1, pp. 11–21.

Rose, P. (2005), "Is there a 'fast track' to achieving education for all?", *International Journal of Educational Development,* Vol. 25/4, pp. 381-394.

Rosenkvist, M.A. (2010), "Using student test results for accountability and improvement: A literature review", *OECD Education Working Papers*, No. 54, OECD Publishing, Paris, http://dx.doi.org/10.1787/5km4htwzbv30-en.

Smith, W.C. (forthcoming),"National testing policies, school practices, and student outcomes: An analysis using data from PISA 2009", *OECD Education Working Papers*, OECD Publishing, Paris.

Smith, W.C. (2014a), "The global transformation toward testing for accountability", *Education Policy Analysis Archives,* Vol. 22/116, http://dx.doi.org/10.14507/epaa.v22.1571.

Smith, W.C. (2014b), *The Global Expansion of the Testing Culture: National Testing Policies and the Reconstruction of Education,* Doctoral dissertation, Pennsylvania State University.

Springer, M.G. (2008). "The influence of an NCLB accountability plan on the distribution of student test score gains", *Economics of Education Review*, Vol. 27/ 5, pp. 556-563.

Volante, L. (2007), "Evaluating test-based accountability perspectives: An international perspective", Paper presented at the Association for Educational Assessment Europe's 8th Annual Conference, 8-10 November 2007, Stockholm, Sweden.

Waterreus, I. (2003), *Lessons in Teacher Pay: Studies on Incentives and the Labour Market for Teachers*, Doctoral dissertation, University of Amsterdam.

West, A. and H. Pennell (2000), "Publishing school examination results in England: Incentives and consequences", *Educational Studies*, Vol. 26/4, pp. 423-436.

West, A., H. Pennell and P. Noden (1998), "School admissions: Increasing equity, accountability and transparency", *British Journal of Educational Studies*, Vol. 46/ 2, pp. 188-200.

Woessman, L. (2007), "International evidence on school competition, autonomy, and accountability: A review", *Peabody Journal of Education*, Vol. 82/ 2-3, pp. 473-497.

Woessman, L. (2004), "The effect heterogeneity of central exams: Evidence from TIMSS", TIMSS-Repeat and PISA, *CESifo Working Paper,* No. 1330.

Wong, M., T.D. Cook, and P.M. Steiner (2009), "No Child Left Behind: An interim evaluation of its effects on learning using two interrupted time series each with its own non-equivalent comparison series", *Northwestern University Institute for Policy Research Working Paper Series,* WP-09-11.

Ysseldyke, J., A. Dennison and R. Nelson (2004), *Large-scale Assessments and Accountability Systems: Positive Consequences for Students with Disabilities*, National Center on Educational Outcomes, University of Minnesota, Minneapolis.

Annex 4.A1: Regression results

Table 4.A1.1 Relationship between testing category (NTP) and student math achievement by school practices

Variable	Model 1: private + controls	Model 2: publicly posted	Model 3: selective admission	Model 4: selective transfer	Model 5: time in Math	Model 6: extra-curricular activities
Evaluative	26.934 (18.411)	23.807 (14.444)	22.030 (15.239)	26.439 (14.683)	13.465 (24.954)	22.611 (14.240)
Formal Sanction/Rewards	42.455* (18.478)	43.070* (18.481)	35.606 (18.341)	48.317** (18.211)	41.112 (21.185)	37.664* (17.196)
*Evaluative * Interaction*		-4.823 (3.112)	3.93 (6.046)	-1.474 (7.101)	0.041 (0.083)	2.295 (3.363)
*Formal Sanction/Reward * Interaction*		-3.057 (4.226)	10.149 (5.679)	-15.076* (6.917)	-0.014 (0.052)	-0.411 (3.111)
*Evaluative * Non-academic extra-curricular*						-3.111 (4.376)
*Formal Sanction/Reward * Non-academic extra-curricular*						4.535 (4.146)
Publicly posted		11.239*** (2.518)				
Admission			13.791** (4.109)			
Transfer				22.713*** (5.895)		
Time in Math					0.122* (0.047)	
Academic extra-curricular						5.779** (2.125)
Non-academic extra-curricular						14.424*** (3.303)
Private	21.391*** (5.085)	21.733*** (5.069)	18.524*** (4.915)	18.416*** (4.925)	21.235*** (4.858)	18.986*** (4.794)
Female	-14.929*** (1.181)	-14.930*** (1.181)	-14.933*** (1.183)	-14.926*** (1.184)	-14.911*** (1.177)	-14.964*** (1.181)
Immigrant status	-10.849* (4.986)	-10.862* (4.987)	-10.912* (4.983)	-10.847* (4.986)	-10.873* (4.987)	-10.907* (4.984)
Home language	-8.189*** (2.181)	-8.172*** (2.188)	-8.153*** (2.186)	-8.195*** (2.192)	-8.177*** (2.198)	-8.123*** (2.176)
SES	15.219*** (1.606)	15.213*** (1.607)	15.178*** (1.613)	15.186*** (1.612)	15.215*** (1.608)	15.138*** (1.618)
_Intercept	459.956 (10.461)	456.393 (10.506)	452.012 (11.027)	451.838 (10.319)	434.023 (14.744)	461.605 (10.153)
Deviance	4683063	4682981	4682752	4682691	4682839	4682555

Note: Three-level random coefficient HLM predicting student math achievement and adjusted for school and student weights. Standard errors in parentheses, interaction effect in italics. Significance levels * p<0.10; ** p<0.05; *** p<0.01 Equation: $Student\ Achievement_{ijk} = \delta_{000} + \delta_{001}$ (NTP) + γ_{00k} (School Type) + γ_{01k} (School Practice) + γ_{02k} (School Practice x NTP) + β_{0jk} (Gender) + β_{1jk}(Student SES) + β_{2jk}(Immigrant Status) + β_{3jk}(Home Language) + $\nu_{00k} + u_{0jk} + e_{ij.}$

Chapter 5.

Making multiple school accountability work

Edith Hooge

TIAS, Tilburg University

The question of how to organise and align different accountability forms and processes has gained relevance as the effects of decentralisation and the introduction of market mechanisms in many OECD countries have become evident. Central governments are still held responsible by the general public for ensuring high quality education, though they play a more limited role as autonomy on the local level has increased. This chapter analyses trends in accountability mechanisms and processes and argues that vertical measures of accountability, that is, regulatory and school performance accountability, can be usefully augmented. The chapter describes how multiple school accountability, that is, horizontal measures involving multiple stakeholders, comes to fruition in different forms and contexts and under which conditions it can flourish. Taking into account the nuanced nature and purposes of education and combining various forms of accountability, multiple school accountability has the potential to enhance the overall education system, policy for reform, and therefore ultimately improve the quality of education.

This chapter is based on earlier work published as: Hooge, E., T. Burns and H. Wilkoszewski (2012), "Looking Beyond the Numbers: Stakeholders and Multiple School Accountability", *OECD Education Working Papers*, No. 85, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k91dl7ct6q6-en.

Introduction

How to hold increasingly autonomous school governing boards and schools accountable for their decisions and performance has become a pressing question for central governments. Over the past three decades, the locus of administrative decision making in education has decentralised in many OECD countries. Local authorities, school governing boards and schools have been allowed a greater degree of freedom in strategy formulation, defining goals and decision making for their education service delivery. This in combination with the enhancement of "customer" (parental) choice and strengthening the quality of the "supply side" by enlarging professional autonomy of teachers and other staff. Yet despite these processes of decentralisation and instituting market mechanisms, central governments are still held responsible by the general public and the media for ensuring high quality education.

Triggered by the results of international benchmarks, such as the OECD Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS), attention is increasingly drawn to the outcomes of educational systems on a national level while goal setting, decision making and publics' appreciation of education quality and outcomes take place on local levels. Consequently, school accountability has become a critical topic, particularly how to align it with accountability for education systems as a whole and how to make it work. This chapter looks at research on existing accountability mechanisms and processes on different levels in education systems and the emergence of new forms of accountability that takes the voices of a diverse set of local and regional stakeholders into account and that can be labelled as multiple school accountability. The chapter is structured as follows. The chapter first discusses the notion of accountability and its implications for education systems and their governance, followed by describing two broader shifts in conceptualising school accountability: from regulatory to performance based accountability and from single to multiple accountability in its most recent form. The chapter discusses examples of multiple school accountability from various countries and gives an outlook of the possibilities, challenges and requirements in adopting a workable concept of multiple school accountability.

Unravelling accountability in education

In research literature, accountability is referred to as "a catchword of the new century" (Herman, 2003:43) but also as "an old and tricky subject" (Barberis, 1998: 451). Although conceived as a "notoriously slippery and multifaceted concept" (Tenbensel et al., 2014: 6), in very general terms, accountability can be defined as processes by which actors are answerable and provide reasons to stakeholders for their actions and/or the actions of their organisation (Acar, Guo and Yang, 2012; Schillemans, 2008; Christensen and Ebrahim, 2006; Pierre and Peters, 2005).

Traditionally, the purpose of accountability in education is legitimation through compliance with laws and regulations. In addition to its legitimation purpose, accountability has also been used as a central vehicle for improvement since the broad school improvement initiatives of the 1990s. This is based on the assumption that holding schools accountable for attaining high standards will, in fact, motivate schools to improve their quality (Geijsel, Krüger and Sleegers, 2010). Today, accounting for, and improvement of, the quality of services provided, in terms of quality of education (effectiveness), value for money (efficiency), equity or access are major purposes of accountability in education, in addition to the legitimation purpose.

Four types of school accountability

In education, two types of accountability mechanisms are commonly used: vertical and horizontal. Vertical accountability is top-down and hierarchical. It enforces compliance with laws and regulation and/or holds schools accountable for the use of resources in relation to the quality of education they provide (efficiency and effectivity).

Horizontal accountability, also identified as "downwards" and "sideways" or "lateral" accountability in the literature, presupposes non-hierarchical relationships. It is directed at how schools and teachers conduct their profession and/or at how schools and teachers inform and involve multiple stakeholders and are accountable to them concerning school's goal setting, strategy formulation, decision-making, implementation and results in terms of quality of educational processes, outputs and outcomes. Each of the two types of accountability is further divided into two subsections: vertical accountability into regulatory school accountability and school performance accountability and horizontal accountability into professional school accountability and multiple school accountability (see Table 5.1 below).

Table 5.1. Types of school accountability

Vertical	**Regulatory school accountability:** Compliance with laws and regulations; focuses on inputs and processes within the school. Mechanism: reporting to higher levels of school authority.
	School performance accountability: Periodic school evaluations. Mechanisms include: 1) standardised student testing, 2) public reporting of school performance, and 3) rewards or sanctions. (Rosendkvist, 2010; Levin, 1974).
Horizontal	**Professional school accountability:** Professional standards for teachers and other educational staff. Mechanisms: credible, useful standards and the creation of professional learning communities (Kim and Lee, 2010; Levitt et al., 2008; Davis, 1991).
	Multiple school accountability: Involving students, parents, communities and other stakeholders in formulating strategies, goal setting, decision-making, and evaluation and appreciation of educational processes, outputs and outcomes (Knutsen and Brower, 2010; De Vijlder et al., 2002; Levin, 1974).

Source: Hooge et al. (2012).

Accountability deficits in education

The question of how to organise and align different accountability forms and processes has gained relevance as the effects of decentralisation and the introduction of market mechanisms in many countries have become evident: central governments play a more limited role as autonomy on the local level has increased. Parallel to this, two accountability deficits are emerging.

Firstly, it appears hard to hold central governments accountable for education policy failures through traditional systems of public accountability such as elections, when central government is only playing an enabling or indirectly controlling role. Conversely, the non-governmental actors directly involved in governing education, such as school governing boards, tend not to be elected officials and, therefore, cannot be held accountable through an electoral process. This brings up the first accountability deficit: the concerns about the legitimacy of schools and school boards that can arise given that they cannot be held directly responsible by the public for their decisions and performance.

Secondly, school performance accountability and the setting of national standards are now commonly used in a majority of OECD countries. Central governments rely heavily on performance accountability systems in order to monitor, control, and steer the quality of education. This enables relatively objective and unambiguous comparison between the performance of schools and educational systems as a whole. The drawback of school performance accountability is that standardised tests in and of themselves cannot reflect the full range of the purposes and goals of schooling such as social skills, moral development, preparation for the labour market, integration, etc., for which schools should be held accountable. The second accountability deficit in education thus concerns the question of how to enhance school performance accountability so that standardised tests could be complemented with other instruments that would assess how well a school or school system is meeting broader standards of education quality and outcomes.

Tensions in accountability

Stakeholder theory provides a useful perspective to study accountability processes, not only because it provides a framework to focus on how schools scan their environment to identify relevant stakeholders as their accountees, but especially because it sheds a light on the inherent tensions in accountability. While accounting for their goals, strategy, decisions, performance and outcomes, schools are faced with different, often mutually conflicting, interests, positions, perspectives and requirements of their stakeholders (Tenbensel et al., 2014; LeRoux, 2009, see also OECD, 2015b).

In education, tensions between so-called vertical accountability and horizontal accountability are likely to be the most prominent as in most countries central government, being held constitutionally responsible for providing quality education, provides funding and sets a legal framework. In general, central governments demand rigorous accounting for resources and lawfulness, often in quantifiable process-output measures. The transaction costs of this vertical accountability in education may be high and can, if not properly aligned, cause tensions with horizontal accountability processes such as meeting professional standards for teachers and educational staff or conceptualising education quality in terms of requirements of parents, institutions for further education or the world of work. For instance, the research project of Kim and Lee (2010) about the impact of competing accountability requirements in non-profit human services agencies (mental health, development disabilities, residential services, community employment and alcohol and drug prevention) reveals that professionals and other agency employees perceive tensions at work as they are increasingly forced to prioritise vertical accountability concerning compliance over professional norms. Applied to education, this means that the often tightly framed and frequently changed report formats and performance standards make teachers and other educational staff feel that documentation is disconnected from their professionals mission. Although agreeing with the professional need to keep track of students' progress and conditions through documentation of educational records, they feel forced to cater more to compliance with vertical accountability mandates. From the perspective of horizontal accountability needs, vertical accountability will be regarded "at best as unwanted distractions and at worst as seriously diluting […] resources and energy" (Tenbensel et al., 2014: 9).

Tensions within horizontal accountability also occur in education. The potential discrepancies between the interests, positions, perspectives and requirements of educational professionals on the one hand and students, parents and community members as lay persons at the other hand are well known and have been brought to light by various

researchers. These discrepancies can lead to tensions between parent- and community-based school governing boards and school leaders, teachers and staff. For instance, Balarin and Lauder (2008) note in UK primary education "a reduced participation of volunteer citizens and members from the parental and wider community in school governance. Existing research points to a widespread divide between the professionals (teachers, head teachers, private business members) and amateurs (parents), which hinders more and better parental involvement" (p. 8). Another example is found in South Africa, where parental involvement in school governance is a relatively new phenomenon. Researchers reveal tensions such as either over-eagerness of parents who want to "run" the schools, or inactivity of parents and educators perceiving participation of parents as beneficial but on the same time expressing concerns about parents overstepping their boundaries. Teachers report to feel uncomfortable with parental involvement in what they define as professional matters and tensions are reported between the values of parents and community members inherent in African traditions and customs and the values of modern school policies and legislations (Brown and Duku, 2008; Heystek, 2006, Grant Lewis and Naidoo, 2004; Van Wyk, 2004). These tensions strengthen when social and demographic differences between the professionals and parents and the community are stronger (Hwang and Powell, 2009).

Tensions between students, parents and community members and other stakeholders of schools also arise in horizontal accountability. Unequal positions of power among these different stakeholders can enable more powerful stakeholders to dominate weaker ones (Brandsen, Oude Vrielink, Schillemans and Van Hout, 2010). This process may take place on a number of levels, including unequal access to decision-making bodies, information and power asymmetries, and the narrowing of the agenda to suit the stronger stakeholders (Fung and Wright, 2001). The possession and use of either professional knowledge or experiential knowledge may be a wedge between schools and their stakeholders, as well as among schools' stakeholders: the position of lay persons (citizens/clients) who merely have experiential knowledge may be weakened as the professionals (dispensing professional knowledge) bond together and strengthen their information exchange and mutual ties (Brandsen et al., 2010).

Accountability shifts in education

The question of how to align different forms of accountability in such a way that accountability deficits can be addressed has gained relevance in many countries. The identification of the different types and forms of school accountability (see Table 5.1) helps explain two recent shifts in accountability in education: (1) the move to complement regulatory school accountability with school performance accountability, and (2) exploring the possibility of moving from singular to multiple school accountability.

Shift in accountability #1: Complementing regulatory with school performance accountability

As laws and regulation are important policy instruments to steer education, regulatory school accountability mechanisms always have been and are still widely used. These primarily include information about students and student characteristics. Less frequent but still common are data on safety issues, curriculum, facilities and grounds, and teacher qualification. The domains with the fewest countries reporting compliance data are related to school finance and governance (OECD, 2011).

In order to balance the greater autonomy granted to schools/school governing boards and the use of market mechanisms, regulatory school accountability has been supplemented with school performance accountability since the 1990s in many OECD member countries. As a general trend, central governments started steering education based on output factors rather than on detailed input factors at this time. Forms of block grant funding were introduced, enabling schools to decide freely how to spend their budget on staff and non-staff costs. In addition, detailed curriculum and classroom organisation prescriptions were replaced by student achievement levels and learning goals to be attained, and market mechanisms were introduced to enhance parental choice and encourage school competition. This shift from input to output steering was accompanied by the introduction of school performance accountability (OECD, 2011; Marks and Nance, 2007; Ladd, 2001). School performance accountability is widespread nowadays in OECD countries, but its frequency and scope vary considerably among and within the countries (see Box 5.1) (OECD, 2012).

Box 5.1. The practice of school performance accountability

Standardised student testing plays an important role in assessing the effectiveness and outcomes of a country's education system. National examinations are standardised tests that have formal consequences for students, such as eligibility to progress to a higher level of education or attain an officially recognised degree. It is most common at upper secondary level (23 of 36 countries reporting) and least present at primary level (4 of 36 countries reporting).

Public reporting of the results of national examinations was mixed in character in the 23 countries where this information was available[1]. Public reporting means that this information is shared with:

- students, school administrators, teachers or parents in the large majority of countries
- media in 11 of 23 countries[2] (OECD, 2012).

Rewarding and sanctioning, the third element of school performance accountability, is much less common. It is done in only four OECD countries: Denmark, Italy, the Netherlands and the United States (OECD, 2011).

Shift in accountability #2: From singular to multiple accountability?

School performance accountability is a good tool for output steering because it enables central governments to steer schools and school governing boards based on their performance. It is a cornerstone of accountability in decentralised educational systems, although as Box 5.1 makes clear, countries have chosen to use it in the way that best suits their individual system. The shift to school performance accountability was an important step in ensuring quality control and effective steering of decentralised systems. However, if governments rely only on school performance accountability in assessing the state of education, essential elements of the quality of education that are not so easy to measure such as socialisation, general knowledge, integration, and personal development may be overlooked. Research has identified a number of unintended effects of school performance accountability (Morris, 2011; Rosenkvist, 2010; Feng, Figlio and Sass, 2010; Resnick, 2006; Kane and Staiger, 2002; Ladd, 2001):

- impoverishing the teaching and learning processes as a result of "teaching to the test"
- narrowing the curriculum in order to focus on those elements that are tested

- emphasising failure instead of learning or improvement if performance accountability lacks positive interventions designed to assist and support low-performing schools
- reducing the quality of staff in schools serving low-performing students.

The higher the stakes are for school leaders and teachers, the more these unintended/undesired effects are likely to occur (Resnick, 2006). Thus, although school performance accountability is a useful tool for central government to monitor quality of student achievement, it is not a cure-all solution when it comes to securing the quality of education in a broad and comprehensive sense. In some OECD countries there has been a move to expand the notion of accountability to a multi-pronged approach that would include the data from school performance measures and augment it with assessment and feedback from other sources (Faubert, 2009; Hooge, van der Sluis and de Vijlder, 2004). These other sources involve elements of multiple accountability and structuring the exchange and relations between relevant stakeholders horizontally.

Horizontal elements in education governance have had a relatively long tradition in a range of OECD countries. School boards or councils comprised of elected, voluntary members have sought to integrate the voices of parents into the governing process, as seen in Austria, Belgium (Wallonia), Germany and lately also in the United States and South Africa (see above). Another example is New Zealand where the local community is strongly involved in school boards' work. Recent policies aim to strengthen these horizontal elements further (OECD, 2015a; Nusche et al., 2012). In some countries, however, notably Denmark, the Netherlands and the United Kingdom, there has been a more recent trend in the public services to move towards more profound multiple school accountability designs.

Defined as a process involving students, parents, communities and other stakeholders in formulating strategies, decision-making, and evaluation for education, multiple school accountability aims to provide: (1) legitimation for the strategy and decision making of the school (is the school doing the right things?), (2) legitimation for the quality of services provided (is the school doing things well?), and (3) improvement of the quality of services provided.

Expanding school performance accountability to encompass a multiple school accountability approach is a potentially promising option for a central government searching for a holistic view of educational quality. In order to think about how and why this could be done, it is useful to distinguish between "process-oriented" and "product-oriented" measurements (such as standardised tests) in school accountability.

Process-oriented measurement is grounded in the idea that school performance can be measured and evaluated from multiple angles and therefore multiple standards and criteria can be used. The use of multidimensional performance measures entails the involvement of multiple sources (quantitative data, qualitative data, narratives, reports, observations) and different actors (inside and outside the school organisation) in measuring a range of processes in schools such as teaching and learning or organisation and leadership. Alternative arrangements such as peer reviews, self-assessments or the involvement of a more diverse set of evaluators (e.g. experts, critical friends, parents) could be adopted to achieve balanced judgments and to take into account factors that are difficult to quantify.

Another way of thinking about a multi-pronged approach is to put schools in a broader context that includes emerging collaboration between organisations in education, welfare, youth care and health and their clients. For instance, communities of practice are created to integrate services and agencies involved in the education and care of children and to encourage the participation of parents, families and communities (Ranson, 2008). A developing practice of extended schools and children's centres have been established in countries such as Denmark, Germany, the Netherlands, Norway, Sweden and the United Kingdom (Cummings, Dyson and Todd, 2011). For example, in the United Kingdom, the Full Service Extended School initiative was introduced in 2003 to provide support for one or more schools in a local education authority area to "provide a comprehensive range of services, including access to health services, adult learning and community activities as well as study support and 8 a.m. to 6 p.m. childcare" (see Cummings et al., 2007).

Developments such as this have consequences for accountability mechanisms: the integration of service delivery from different institutions and organisations requires accountability mechanisms with an accumulated and integrated character. Acar and colleagues (2012) have carried out a research project on how to achieve accountability in collaborative forms of governance such as multi-organisational partnerships. They conducted a field study on the views of practitioners from voluntary partnerships that were formed between K-12 public schools and private and non-profit organisations such as community groups, businesses, universities and government agencies in the United States, all with the purpose of promoting student success. The practitioners from these partnerships seemed to be more concerned with the "for what" dimension of accountability e.g. enhancing student achievement and development, meeting goals and objectives of the partnership, preparing future workforce, reducing absenteeism, providing adult role models and so on, than with the "to whom" accountability question. With respect to the "to whom" question, they indicated to feel accountable primarily to students (the beneficiaries of the partnerships) and also to partners, businesses and schools. To a far lesser extent, they felt accountable to partnership offices, school districts, partnership boards of trustees or legislature. These findings show that practitioners of multi-organisational partnerships hold more client-based and results-oriented views of accountability.

In short, in order to reduce unintended effects of school performance accountability, interest in multiple school accountability has grown in the last ten years (Morris, 2011; Faubert, 2009; Ozga, 2009). A form of horizontal accountability (see Table 5.1), multiple school accountability means that schools are accountable to students and their parents, to members of the community, and to the community as a whole for multiple aspects of schooling, based on various information sources (Biesta, 2004; Levin, 1974). Multiple school accountability aims to increase legitimacy and trust from the local community through the processes of learning and feedback that it receives (Hooge and Helderman, 2008; De Vijlder and Westerhuis, 2002). It requires that schools work closely with different stakeholders, supporters and constituents in their environment in order to:

- help them learn about their rights and duties, requirements, desires and expectations concerning education
- establish a relationship (by negotiating, collaborating and/or involving them)
- obtain support for school policies, strategy, decisions and practices, and
- be held accountable by them.

The emergence of multiple school accountability

Multiple school accountability comes to fruition in different forms and contexts. Examples of multiple school accountability at work are found in:

Denmark

In Denmark for instance, multiple school accountability is found in a basic form. The governance of primary and lower secondary schools is divided between two bodies: the local or town council (*Kommunalbestyrelse*) and the school council (*Skolebestyrelse*). The first opens and closes schools, hires and fires teachers, and administers the budget; the latter advises the local or town council with regard to the design of curricula and the activities of the school. It is comprised of five to seven elected representatives of the parents; joint sessions of the town and school councils are also attended by the head of school, teachers and students. The head of school is accountable to both bodies (Stückler, 2005).

England and Wales

In England and Wales, processes of multiple school accountability are more developed. Here, every school has a school governing body that is comprised of the head of school, elected representatives of parents, teachers and non-teaching school staff, the local education authority, as well as local political representatives. The body is responsible for general administration (including budget) and hiring and firing of teachers and heads of schools (Stückler, 2005).

Scotland

In Scotland, the national assessment development programme "Assessment is for Learning" (AifL) focused roughly between 2002 and 2012 on aligning "assessment for learning" and "assessment for accountability". AifL implies multiple accountability, which in the programme is labelled as "intelligent accountability". Hutchington and Young (2012) assert in their evaluation study of AifL (2012) that specific reporting arrangements are a condition for putting multiple or intelligent accountability into practice. In the AifL programme a series of "Open Space" events for parents were launched to discover parents' expectations for reporting: "Contrary to expectations, parents' focus was clearly on 'learning for life' and on knowing what their children were learning and their strengths and development needs, so that they could support them in partnership with schools. Parents' feedback refers explicitly to the desire for assessment to support learning, not more tests" (Hutchington and Young, 2012: 66). Subsequently, an explanatory leaflet for parents about AifL was published to support schools in their communication with parents and a number of learning communities were supported to explore ways of recognising and reporting the whole range of students' achievements.

The Netherlands

In the Netherlands there is a strong movement towards multiple school accountability. All Dutch education governance codes require schools to identify relevant stakeholders and involve them in strategy formulation, goal setting, decision-making, and evaluation and appreciation of educational processes, outputs and outcomes, referred to as "conducting a horizontal dialogue". The pull of these governance codes is strengthening as the national organisations of school governing boards set compliance of these governance codes as a membership requirement. Another interesting initiative in the Netherlands is the project called Windows for Accountability (*Vensters voor*

verantwoording), piloted from 2007 by the *VO-raad*, the organisation of secondary school governing boards. In 2010, Windows for Accountability was rolled out as a nationwide project. Although it is a voluntary service, currently it is being used by 94% of Dutch secondary schools. Website based, this project posts information on the organisation and quality of Dutch secondary schools in a simple and standardised way. The website is accessible to the public and contains quantitative data from standardised tests and assessments as well as explanatory comments from schools on their teaching practice, learning outcomes, the quality of the teachers, school climate, etc. The *PO-raad*, the organisation of primary school governing boards, joined this project in 2012 and strives for 100% national coverage in 2015. In 2014, information about 80% of Dutch primary schools is available.

California, USA

In 2013, a community-based approach that creates Local Control Funding Formula (LCFF) processes for school finance was introduced in California, USA. Vasquez Heilig et al. (2014) carried out a statutory analysis of LCFF and find that all districts and charters receiving funding under the LCFF are required to develop a local accountability plan (LCAP) that must include district or charter specific goals and priorities, addressing state and local priorities. Quantitative as well as qualitative measures may be formulated to gauge the path and progress toward the goals and the school or governing board is required to solicit input from all interested parties when developing its LCAP: teachers, principals, administrators, education service professionals, local bargaining units, parents and students. Furthermore, community control is established by requiring that the LCAP is reviewed by advisory committees, subject to public comment, and heard in at least two public hearings.

Vocational Education and Training

Co-operation between Vocational Education and Training (VET)-institutions and the labour market are illustrative for multiple school accountability at work. In VET systems, connectivity with the world of work and with society is considered very important. In order to bridge and align education, training and work, VET institutions are assumed to take into account perceptions and convictions of employers about the purposes, goals, content, pedagogy and quality of Vocational Education and Training (VET). There are multiple ways for VET institutions to do this, and different examples of formal and informal feedback mechanisms of VET labour market co-operation throughout Europe have been identified, see Table 5.2 below.

Table 5.2. Examples of formal and informal feedback mechanisms of VET labour market co-operation

Formal feedback mechanisms	Informal feedback mechanisms
• Sector skills councils, e.g. the co-operation between employer-led organisations and the Commission for Employment and Skills in England. • Trade committees, e.g. Denmark, bipartite arrangements of employer associations and trade unions supporting new VET, adjustments or closing of outdated programmes. • Advisory boards on apprenticeships, e.g. Austria, where social partners are involved in the process of developing/renewing occupational profiles. • Managing boards and expert committees for VET, e.g. Bulgaria, where state and social partners cooperate in the development and renewal of school-based VET curricula commissions to be found in most countries examined.	• local school boards • professional internships • exchange programs • dual systems, work-based training • alumni networks • career fairs • projects in companies • school at work initiatives (in-company learning in co-operation with schools) • work at school initiatives (experienced professionals provide supervision and professional skills training in school)

Source: Cedefop (2013: 26)

Outside the field of education

Examples from outside the field of education are Poland, the United Kingdom and again, the Netherlands. These countries have taken steps towards introducing multiple accountability systems in the public services. A research project on 82 Dutch public agencies shows that horizontal accountability processes work to foster richness of information and new insights in organisational learning. The reflective dialogues with stakeholders aim to improve rather than judge; judgement is not used as a means of control but rather for advising and giving operational lessons. As such, there is no short-term accountability pressure in terms of presenting immediate results and actors in these processes are able to take a strategic, longer-term perspective. The judgements described appear to be less driven by short-term political considerations but, rather, are more concerned with the quality of service over time (Schillemans, 2008; for related research on multiple accountability in the Dutch housing sector see SEV, 2006).

Within the field of corporate governance, some countries have also moved to systems of multiple accountability. In the United States and the United Kingdom for example, so called "Say-on-Pay" regulations have enabled shareholders to express their voice by voting on the pay policy of the company's executive officers. This vote does not focus on pay itself, but rather on the balance between compensation and performance of the corporation. Proposals that pass the majority threshold are not necessarily binding for the executive board. However, they do exert pressure on the board members to reflect on executive pay and its efficacy to deliver performance. Recent research has shown that Say-on-Pay appears to lead to large increases in market value, profitability and long-term performance in large corporations (Cuñat, Gine and Guadalupe, 2013).

Multiple accountability is thus an intriguing option for governments interested in augmenting the scope and feedback loops of their accountability systems. Yet, as multiple accountability still is a fairly novel approach, the question remains: Does it really work? Initial reports are mixed: there is great appreciation for the process and a broader range of stakeholder voices. However, ministries report a reluctance to rely too heavily on information generated by multiple accountability mechanisms due to doubts about its reliability and the risk of information overload. On the basis of this, central government is advised to discuss the purposes and use of multiple accountability mechanisms with the institutions and to balance the opportunities (information to learn, improve, steer, and formulate policies) with the risks (e.g. information overload) (Dutch Court of Audit, 2011). If multiple accountability mechanisms are indeed perceived as complementary to vertical accountability mechanisms, central government has to clarify how and to what extent this is undertaken. It is also essential to manage the expectations of the organisations and individuals involved (Brandsen et al., 2010). Agreement about the nature and extent of extra information is needed as is more research on how multiple accountability works and its effects. The last section of this paper focuses on the practical side of multiple accountability and how it could work in schools.

How to make multiple school accountability work in education

Horizontal accountability is a worthwhile but difficult endeavour. The centrality and "pull" of vertical (hierarchical) accountability towards governments and inspectorates is prominent and can crowd out professional and multiple accountability processes (LeRoux, 2009; Tenbensel et al., 2014). At its best, multiple school accountability is a process where, having gathered real insight into school's strengths and weaknesses, a school meets with its accountees to conduct a fruitful dialogue about the school's

decisions and performance in relation to the perceptions, expectations and judgments of different stakeholders. For this to happen, the relevant stakeholders need to be identified, and in some cases motivated and/or trained. Schools themselves need to build capacity in terms of leadership for multiple school accountability processes, and also in terms of the ability to interpret and correctly use data from school performance accountability (e.g. assessment results) (see also OECD, 2015b). Here the various processes involved in each of these activities are briefly described and examples of activities in this domain are provided.

Identifying stakeholders

Which organisations, groups or persons are important for the legitimation of the school's strategy, decision-making, and the quality of the service delivery? And which parties are in a position to evaluate and give valuable feedback in order to improve the quality of education? With respect to multiple accountability processes, Hooge and Helderman (2008) distinguish four different categories of stakeholders: primary, internal, vertical, and horizontal.

In education, parents and students are the primary stakeholders. Teachers and other educational and non-educational staff are internal stakeholders with a clear interest in the success of the school. At slightly more distance, governments and organisations formally operating on behalf of government (such as inspectorates or municipalities) operate as vertical stakeholders. Finally, all other organisations, groups or persons in the school's environment with some level of interest in the school are horizontal stakeholders.

Engaging parents in multiple accountability processes might thus help improve student performance. Greater parental engagement can be fostered by clarifying ways in which parents can contribute and participate, by ensuring that the purposes of parent engagement are explicit; by providing training for parents to play an advocacy role, by strong school leadership, and, most importantly, by instituting a decision-making framework that provides parents with real influence and voice in decision making (Caldwell, 2012; Leithwood, 2009).

Building stakeholders' capacity

It does not always occur to many stakeholders to act as an accountee towards a school. If a school fails to pay attention to the knowledge, motivation or positions of stakeholders as potential accountees, then valuable but weaker stakeholders risk being excluded. This lessens the quality of multiple accountability processes.

Although multiple accountability processes are of a non-hierarchical character, this does not mean that the relationship between the school and horizontal accountees is equal in every respect. In some instances, the stakeholder has the same level of organisation, knowledge and involvement as the school, but in other situations, the stakeholder will have less knowledge and involvement than the school itself. This may mean that schools are sometimes better motivated and equipped to enter into the accountability processes with their stakeholders than vice versa. Schools need to approach stakeholders in proper proportion. Some stakeholders need help with acquiring knowledge and organising their involvement as an accountee. While it might be easy or tempting to "score" against certain of those stakeholders, this raises the question of whose interest this would ultimately serve. And, of course, the same applies in the reverse situation.

Potential pitfalls

Being engaged in multiple accountability processes may demand too much from parents and involved members of the community. Often they lack the knowledge, time, patience and wisdom that they are expected to dispense, or, to acquire in short order through training, or they may be unaware of the issue or too consumed with private life. Real frustration arises when parents and involved members of the community perceive that the rules of the game are dictated by the school and the communication is a "one-way street" rather than a truly free and open dialogue (Leithwood, 2009). In reaction to this, parents and community members may choose not to engage, which sometimes leads schools to incorrectly conclude that they are satisfied with schools' delivery of services. This is a lost opportunity to both engage important stakeholders and improve service and achievement.

Consultation and participation fatigue may also be a pitfall of multiple accountability. Brandsen et al. (2010) conclude that multiple accountability increases the accountability pressure since it complements traditional vertical accountability rather than substitutes for it. "Many of the organisations we examined indicated that they felt burdened by an increasing amount of paperwork [...] the fatigue of staff members was mirrored by stakeholders, especially individual clients, who showed increasing disinterest in being consulted and involved" (p. 17).

Building schools' capacity

The work of school leaders is crucial to building school capacity for multiple accountability because accountability processes are nested in beliefs, experiences and practices in schools. It requires school leaders who are willing and able to empower staff, and in turn, to involve and share responsibility with parents and other interested members of the local community. It also requires school leaders who are willing to be held accountable by them (Leithwood, 2001).

Apart from leadership, the capacity to handle data is a key element of school capacity building with respect to accountability (Ozga, 2009). Masses of data are available through assessment and monitoring systems, indicators of effectiveness, targets, inspection and review programs. Methods for accessing information and, consequently, analysing and interpreting it, are not self-evident in schools. As early as the 1970s, many of the relevant data were not available for schools or at least not in a form which could be easily used (Levin, 1974). Apart from the lack of availability and feasibility of data, until recently there was often a gap between the interests expressed in data and the actual use of data. Schools need the capability to transform data into knowledge appropriate for multiple accountability purposes. This requires proper school self-evaluation: obtaining real insights into the quality and processes of schools that are relevant to the practice of accountability on multiple fronts.

Coping with data requires that educators themselves become experts in interpreting data and transforming it into knowledge. Earl and Katz (2002; 2006, cited in Geijsel et al., 2010: 62), point to three capacities that school leaders need in order to work in a data-rich world:

- *Develop an inquiry habit of mind.* Leaders need to reserve judgment and have a tolerance for ambiguity, to value deep understanding, take a range of perspectives and systematically pose increasingly focused questions.

- *Become data literate.* Leaders must to be aware of how different data are needed for different purposes; they need to be able to evaluate data, recognising sound and unsound data, to be knowledgeable about statistical and measurement concepts, to recognise other kinds of data (not only numbers, but also opinions, anecdotes, observations), to make interpretation paramount (instead of using data for quick fixes), and to pay attention to reporting to different audiences.
- *Create a culture of inquiry.* Leaders need to involve others in interpreting and engaging with the data, to stimulate an internal sense of urgency (re-focusing the agenda), to make time for data interpretation and for coming to collective meaning and commitment, and to use critical friends.

Schools can also take advantage of the potential of Internet-based technologies to address accountability, see Box 5.2.

Box 5.2. Accountability online

Internet-based technologies provide stakeholders with an increasing ability and interest to gain access to information about schooling and schools they deem important, and on the supply side, these technologies enable schools to disclose information. Two purposes of accountability online can be distinguished.

1. Disclosure. This can be achieved by posting content on the website such as mission statement, history, vision, plan, values and goals, budgeting materials, reporting on using financial resources and compliance related documents as well as data and information about education outputs, outcomes and community impacts.
2. Dialogue. This can be done by tapping stakeholders' needs, preferences and demands, enabling stakeholders to have some degree of say in decision-making and policies, brainstorming and problem-solving through feedback forms, discussions lists, bulletin boards, collaborative wiki's, online surveying and polling tools, tagging and social bookmarking projects, webinars.
3. To examine the extent to which non-profit organisations adopt Internet-based accountability, Saxton and Guo (2014) analysed data of 117 US community foundations. The majority had the most basic "contact-us", feedback or ask a question on their websites, whereas only 7% used higher-level mechanisms such as online stakeholder survey, interactive message forum or an online needs assessment. Severe underutilisation of the technology was thus concluded. In other words, the opportunity to use the Internet to engage stakeholders can be greatly improved.

Source: Saxton and Guo (2014).

The practice of multiple accountability has yet to come to fruition in education, and the amount of available research on this topic is modest. Based on theory and experience from other sectors however, some lessons can be learned to make multiple school accountability work:

- It is important to identify the right stakeholders. The process of stakeholder identification can be heavily influenced by "stakeholder salience", that is, the ability of stakeholders to attract schools' attention, depending on their power, legitimacy and urgency vis-à-vis the school (Mitchell, Agle and Wood, 1997). In order to ensure that the identification of stakeholders is not limited to those most salient, schools must make efforts to involve less powerful or inactive

stakeholders. Being less powerful or inactive does not mean that these stakeholders are not relevant to the school. On the contrary, these are often the very stakeholders for whom the school aims to add value; therefore, schools need them.

- Build stakeholder capacity. This is particularly important while establishing accountability relationships with weaker stakeholders who might not have the requisite knowledge and language to play the role of an accountee and, therefore, may inadvertently be excluded in accountability processes. Avoiding apathy and "consultation fatigue" is key because they weaken the effectiveness of the process, and ultimately the strength of this approach is determined by its weakest accountees. Schools can involve and activate their stakeholders by being inviting, by structuring participation and accountability processes, and by motivating and empowering them.
- Self-evaluation that provides real insight into schools' quality and processes is needed to make multiple accountability work. Proper school self-evaluation requires "assessment literacy" (Fullan, 2007; Hutchinson and Young, 2011) from school leaders as well as from teachers and other professional staff. The work of school leaders is crucial here: they must empower staff to be involved and open to parents and members of the local community and to be held accountable by them, and they must create the effective environments by building bridges between teachers and educational staff and external accountability demands. Autonomy and a (governance) environment that provides support foster this work of school leaders.

A warning: multiple school accountability is not a panacea. It would be simplistic to rely solely on this one concept to solve local-level accountability issues because this might also lead to unintended or undesired effects. First, school leaders and teachers can use defensive reasoning and be wary of scrutiny and interference from the wide range of stakeholders involved in multiple accountability. They may wish to avoid accounting for their decisions, practices and outcomes, and consequently, give accountability relationships a symbolic or fake character. Second, since multiple school accountability relies heavily on the perceptions and experiences of school stakeholders, there also is a risk of only mapping stakeholders' (dis)satisfaction, coloured by social desirability and/or "myths" concerning the image of the school. Third, the use of market mechanisms such as school competitiveness and parental choice in education can be disincentives for making multiple school accountability truly work because sometimes too much transparency concerning the weaknesses of a school may threaten a school's image.

Conclusions

This chapter addresses the following question: How can schools and school boards be held accountable to the public for their decisions and performance? In decentralised educational systems, it is no longer enough that autonomous schools/school boards are held accountable only for compliance with input factors, as required by law and regulation; they must also meet the required quality standards for service delivery. School performance accountability is a good tool for this as it enables central governments to steer schools and school governing boards on the basis of their performance. In this respect, school performance accountability is a cornerstone of accountability in decentralised educational systems.

However, school performance accountability does not allow for assessing such elements as socialisation, general knowledge, integration and personal development. It also does not look at building local confidence and legitimacy. Recently, in some countries there has been a trend to move towards multiple school accountability. Multiple school accountability takes into account different stakeholders' varying perceptions of the quality, effectiveness and efficiency of schooling. It can complement school performance accountability by looking beyond the numbers and also defining schooling in professional and democratic terms.

Last, but not least, it aims to foster transparency in the system by opening it up to public scrutiny. In practice, the government is advised to discuss the purposes and use of multiple accountability mechanisms and to balance the opportunities (information to learn from and to use to improve, steer and formulate policies) with the risks (information overload). If multiple school accountability mechanisms are indeed perceived as complementary to vertical accountability mechanisms, central government has to align it through agreeing with school organisations on how and to what extent multiple accountability mechanisms are to be introduced and used.

There have been numerous shifts in accountability practice and research over the last few decades. Accountability issues are a central priority for OECD countries and one of the hottest debates currently going. This chapter has sought to argue that vertical measures of accountability, that is, regulatory and school performance accountability, can be usefully augmented by horizontal measures involving multiple stakeholders. This combination aims to build an efficient and effective accountability system that takes into account the nuanced nature and purposes of education. Combining various forms of accountability will help to improve the overall education system, policy for reform, and therefore ultimately improve the quality of education.

Notes

1. Australia, the Czech Republic, Denmark, England (United Kingdom), Estonia, Finland, France, Germany, Hungary, Iceland, Ireland, Israel, Italy, Luxembourg, the Netherlands, Norway, Poland, Portugal, Scotland, the Slovak Republic, Slovenia, Spain, the United States, Indonesia and the Russian Federation.
2. Denmark, Estonia, France, Norway, Poland, Portugal, Scotland and the Russian Federation.

References

Acar, M., C. Guo, and K. Yang (2012), "Accountability in voluntary partnerships: To whom and for what?", *Public Organisation Review*, Vol. 12/2, pp. 157-174.

Barberis, P. (1998), "The new public management and a new accountability", *Public Administration*, 76/3, pp. 451-470.

Balarin, M. and H. Lauder (2008), "The governance and administration of English primary education", Interim report, University of Cambridge.

Ball, A. and S.P. Osborne, *Social Accounting and Public Management*, Routledge, London.

Biesta, G.J.J. (2004), "Education, accountability, and the ethical demand: Can the democratic potential of accountability be regained?", *Educational Theory*, 54/3, pp. 233-250.

Brandsen, T. et al. (2010), "Non-profit organisations, democratisation and new forms of accountability: a preliminary evaluation", in P. Osborne and A. Ball (eds.), *Social Accounting and Public Management*, Routledge, London, pp. 90-102.

Caldwell, B. (2012), "Review of related literature for the evaluation of empowering local schools", Council for Educational Research, Australia.

Cedefop (2013), "Renewing VET provision: Understanding feedback mechanisms between initial VET and the labour market", *Research Paper no.37,* European Centre for the Development of Vocational Training (Cedefop), Luxembourg.

Cummings, C., A .Dyson and L. Todd (2011), *Beyond the School Gates: Can Full Service and Extended Schools Overcome Disadvantage*?, Routledge, London.

Cuñat, V., M. Gine and M. Guadalupe (2013), "Say pays! Shareholder voice and firm performance", *Upjohn Institute Working Paper 13-192*, Upjohn Institute for Employment Research. Kalamazoo, MI, http://dx.doi.org/10.17848/wp13-192.

De Vijlder, F.J and A. Westerhuis (2002), "Meervoudige publieke verantwoording: Een aanzet tot conceptualisering en een verkenning van de praktijk" [Multiple public accountability: conceptualisation and exploring the practice], Max Goote Kenniscentrum, Amsterdam.

Dutch Court of Audit (2011), "Gebruik van horizontale verantwoordingsinformatie" [The Use of Horizontal Accounting Information], RWT-verkenningen deel 3, Algemene Rekenkamer, The Hague.

Faubert, V. (2009), "School evaluation: Current practices in OECD countries and a literature review", *OECD Education Working Papers*, No. 42, OECD Publishing, Paris, http://dx.doi.org/10.1787/218816547156.

Feng, L., D. Figlio and T. Sass (2010), "School accountability and teacher mobility", *Working Paper 47,* Calder Urban Institute.

Fullan, M. (2007), "The new meaning of educational change", Teachers College Press, New York.

Fung, A. and E.O. Wright (2001), "Deepening democracy: Innovations in empowered participatory governance", *Politics and Society*, Vol. 2/1, pp. 5-41.

Geijsel, F.P., M.L. Krüger, and P.J.C. Sleegers (2010), "Data feedback for school improvement: The role of researchers and school leaders", *The Australian Educational Researcher*, Vol. 37/2, pp. 59-75.

Herman, J. (2003). "2003 annual meeting theme and highlights – accountability for educational quality: shared responsibility", *Educational Researcher*, 32/1, p. 43.

Hooge, E.H. and J.K. Helderman (2008), "Klant en overheid koning. Over toezicht op meervoudige verantwoording door maatschappelijke ondernemingen" [Client and Government King? About Monitoring Internally Horizontal Accountability by Dutch Societal Entrepreneurs], *Bestuurskunde*, Vol. 17/3, pp. 95-104.

Hooge, E.H., M.E. van der Sluis and F. J. de Vijlder (2004), "Stakeholders in beeld. Over instellingen voor beroepsonderwijs en hun stakeholders en over methoden om stakeholders te identificeren en te postioneren" [Visualising Stakeholders. About Educational Institutions and Their Stakeholders], Max Goote Kenniscentrum, Amsterdam.

Hutchinson, C. and M. Young (2011), "Assessment for learning in the accountability era: Empirical evidence from Scotland", *Studies in Educational Evaluation*, Vol. 37/1, pp. 62-70.

Hwang, H. and W.W. Powell (2009), "The rationalisation of charity: The influences of professionalism in the nonprofit Sector", *Administrative Science Quarterly*, Vol. 54/2, pp. 268-298.

Kane, J.K. and D. O. Staiger (2002), "The promise and pitfalls of using imprecise school accountability measures', *Journal of Economic Perspectives*, Vol. 16/4, pp. 91-114.

Kim, S.E. and J.W. Lee (2010), "Impact of competing accountability requirements on perceived work performance", *The American Review of Public Administration*, Vol. 40/1, pp. 100-118.

Knusten, W.L. and R.S. Brower (2010), "Managing expressive and instrumental accountabilities in nonprofit and voluntary organisations: A qualitative investigation, *Non-profit and Voluntary Sector Quarterly*, Vol. 39/, pp. 588-610.

Ladd, H.F. (2001), "School-based educational accountability systems: The promise and the pitfalls", *National Tax Journal*, Vol. 54/2, pp. 385-400.

Leithwood, K. (2009), "Four key policy questions about parents engagement recommendations from the evidence", in Deslandes, R. (ed.), *International Perspectives on Contexts, Communities and Evaluated Innovative Practices*, Family-School-Community Partnerships, Routledge, London, pp. 8-20.

Leithwood, K. (2001), "School leadership in the context of accountability policies", *Leadership in Education*, Vol. 4/3, pp. 217-235.

Leroux, K. (2009), "Managing stakeholder demands: Balancing responsiveness to clients and funding agents in non-profit social service organisations", *Administration and Society*, Vol. 41/2, pp. 158-184.

Levin, H.M. (1974), "A conceptual framework for accountability in education", *The School Review*, Vol. 82/3, pp. 363-391.

Marks, H.M. and J.P. Nance (2007), "Contexts of accountability under systemic reform: Implications for principal influence on instruction and supervision", *Educational Administration Quarterly*, Vol. 43/1, pp. 3-37.

Mitchell, R.K., B.R. Agle and D.J. Wood (1997), "Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts", *Academy of Management Review*, Vol. 22/4, pp. 853-886.

Morris, A. (2011), "Student standardised testing: current practices in OECD countries and a literature review", *OECD Education Working Papers*, No. 65, OECD Publishing, Paris, http://dx.doi.org/10.1787/5kg3rp9qbnr6-en.

Nusche, D. et al. (2012), "School education in New Zealand", *OECD Reviews of Evaluation and Assessment in Education: New Zealand 2011*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264116917-5-en.

OECD (2015a), *Education Policy Outlook 2015*: *Making Reforms Happen*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264225442-en.

OECD (2015b), *Schooling Redesigned: Towards Innovative Learning Systems*, Educational Research and Innovation, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264245914-en.

OECD (2012), *Education at a Glance 2012: OECD Indicators*, OECD Publishing, Paris, http://dx.doi.org/10.1787/eag-2012-en.

OECD (2011), *Education at a Glance 2011: OECD Indicators*, OECD Publishing, Paris, http://dx.doi.org/10.1787/eag-2011-en.

Oostdam, R. J. and E.H. Hooge (2012), "Making the difference with active parenting; forming educational partnerships between parents and schools", *European Journal of Psychology of Education*, Vol. 28/2, pp. 337-351.

Ozga, J. (2009), "Governing education through data in England: From regulation to self-evaluation", *Journal of Education Policy*, Vol. 24/2, pp. 149-162.

Pierre, J. and B.G. Peters (2005), *Governing Complex Societies*, *Trajectories and Scenarios*, Palgrave MacMillan, London.

Ranson, S. (2008), "The changing governance of education", *Educational Management Administration & Leadership*, Vol. 36/2, pp. 201-219.

Resnick, L.B. (2006), "Making accountability really count", *Educational Measurement: Issues and Practice*, Vol. 25/1, pp. 33-37.

Rosenkvist, M.A. (2010), "Using student test results for accountability and improvement: A literature review", *OECD Education Working Papers*, No. 54, OECD Publishing, Paris, http://dx.doi.org/10.1787/5km4htwzbv30-en.

Saxton, G.D. and Ç. Guo (2011), "Accountability online: understanding the web-based accountability practices of nonprofit organisations, *Nonprofit and Voluntary Sector Quarterly,* Vol. 40/ 2, pp. 270-295.

Schillemans, T. (2008), "Accountability in the shadow of hierarchy: The horizontal accountability of agencies", *Public Organisation Review*, Vol. 8/2, pp. 175-194.

SEV (2006), "Belanghouders identificatie tool" [Tool to identify stakeholders], Housing experiments steering group, Rotterdam.

Stückler, A. (2005), "Die Wirksamkeit der Schulmanagementausbildung auf die Tätigkeit der Schulleiter – Eine Evaluationsstudie an Kärntens Pflichtschulen" [The effectiveness of the school management training to the business of the headmaster: An evaluation study of Carinthia compulsory schools], Doctoral dissertation, Alpen-Adria-Universität Klagenfurt, www.seb.ph-kaernten.ac.at/downloads/dissertation2005-05-08.pdf.

Tenbensel, T., J. Dwyer and J. Lavoie (2014), "How not to kill the golden goose: Reconceptualising accountability environments of third-sector Organisations", *Public Management Review*, Vol. 16/7, pp. 925-944.

Vasquez Heilig, J. et al. (2014), "Community-based school finance and accountability: A new era for local control in education policy?", *Urban Education*, vol. 49/8, pp. 871-894.

PART 3.

CAPACITY AND THE USE OF KNOWLEDGE

Chapter 6.

Complexity in a bureaucratic-federalist education system

Lorenz Lassnigg

Institute for Advanced Studies, Austria

On the case of Austria the chapter explores some main issues of complexity in centralised systems. In a first part, the chapter describes that while most sources of complexity in centralised systems generally add to those found in decentralised systems, the degree of centralization (or decentralisation) should not be perceived as dichotomy as crucial for a systems structural complexity is its specific setup.

Building on this, the chapter describes how the tensions between policy and politics as basic dimensions of governance and policy making are greater in bureaucratic-federalist systems such as Austria due to their structurally complex setup. An important aspect of the whole interrelations in a centralised system lies in the fact that much part of the complexity is hidden behind the existing formal regulations that superficially seem to "rationalise" practices, however, might create a substantial gap between formal structures and informal practices.

This chapter is based on earlier work published as: Lassnigg, L. (2009), *Policy Learning and Outcome Orientation in Austrian VET – Theoretical and Empirical Explorations*, in F. K. Oser et al. (eds.), *VET Boost: Towards a Theory of Professional Competencies*, *Essays in Honor of Frank Achtenhagen*, Sense Publishers, Rotterdam/Chinese Taipei, pp. 385-408.

Lassnigg, L. et al. (2007), "Ökonomische Bewertung der Struktur und Effizienz des österreichischen Bildungswesens und seiner Verwaltung" [Economic Assessment of the Structure and Efficiency of the Austrian School System and its Administration], IHS research report commissioned by Bundesministerium für Unterricht, Kunst und Kultur (BMUKK), www.equi.at/dateien/ihs_oekbew.pdf.

Introduction

This chapter has two purposes, first to relate previous work about the governance of the Austrian education system in an explicit and systematic way to understandings or concepts of complexity in policy making, and second to explore the implications of the distinction between centralised and decentralised governance systems for complexity in policy making. Complexity is an ambiguous and highly unwelcome issue in policy making, as it makes things and processes complicated. The author remembers several occasions in policy discourses where participants demanded from research to simplify the objects dealt with; thus complexity should be somehow defined away by simplifying things through observation and presentation. There is even a formalised expression for this: KISS – Keep it Simple (and) Stupid.

However, these wishes mostly imply a misunderstanding of what complexity means: to reduce statements about complicated things to their main factors or traits. The misunderstanding is first confusing complex and complicated issues, and second to confuse complexity at the level of practices, processes or relationships with complicated descriptions at the level of observation. So the hope is to reduce the complexity at the practice level through simplified descriptions: complicated phenomena created by many interacting variables which are difficult to oversee should be first analysed and then reduced to the main ones in understanding (e.g. by modelling). The meaning of complexity does not necessarily involve many variables, but it involves unpredictable dynamics (which can be created already by few variables), which is given in real practices or processes. In this meaning the reduction of complexity cannot be done by observation but must be handled in practice.

Thus relating concepts of complexity to an existing governance system implies epistemological decisions and a good deal of interpretation: it is always possible to seek for more and better variables to predict the unpredictable, and it can be predicted that this will always be ongoing in a traditional perspective; to follow this path will of course also add information and knowledge to the understanding of existing systems (at least if it shows that certain variables or constellations do not explain anything). The complexity perspective means to take another (constructivist systemic) path of looking at a system from different assumptions which shift from complicacy to complexity and thus take unpredictability at face value and look at which kinds of solutions might follow from that perspective.

The different perspectives can be illustrated by the distinction between the sources of increasing complexity in decentralised systems which mainly consist of adding "variables" and their properties to a given state (more actors with more weight), and complexity in centralised systems which has to do with constellations between given factors, which might be only few. In the following analysis and interpretation in particular two phenomena are used for explanation, the first pertaining to the interaction between three different governance mechanisms (bureaucracy, federalism and corporatism) which might in fact involve the same actors, and second, the differences between the designed working of devised structures "on paper" (e.g. by regulations, or organigrams) and their real enactment if regulations are "filled with life".

As will be shown, the seemingly abstract problem of the interaction between these two phenomena poses very real questions of current reform in the Austrian system of educational governance: which "real" consequences might follow from a simplification of the distribution of responsibilities between the governance levels "on paper", when the changed structures will be "filled" with the given actors and their practices, power

relations and the like. This also poses questions about the use and impact of knowledge in governance, as a prediction of the outcomes of a reform would involve also the interaction of different kinds of knowledge, the practical knowledge owned by the actors, and formal knowledge gained at the level of observation and analysis (e.g. by advisors or evaluators), whereby the different kinds of knowledge of the different actors must be conceived as an element in the power play also. Consequently, different strategies of the provision and use of knowledge result from a traditional complicacy perspective vs. a complexity perspective, with a transportation of information about "evidence" from the observers to the actors intended from the former, and a push towards reflexivity and active knowledge production among the actors from the latter perspective (where the observers are conceived of as a certain kind of involved actors).

The chapter analyses first the sources of complexity in a centralised system which are different from those so far handled in decentralised systems in the GCES project, and illustrates this by taking the Austrian system as a specific case which seems to have a quite particular structure. Secondly, an interpretation of the political dynamics related to centralised systems is given by using the distinction between policy and politics as a main explanatory device. As centralised systems constitute different actor constellations from decentralised systems, different relationships between policy and politics might arise in the different kinds of structures, with centralised systems on the one hand giving more weight to the politics dimension, however, being less able to handle the policy issues because of the structural complexities. That is, that education might be endemically pushed more strongly up to the level of "hot" government politics in centralised systems, without them having good conditions to handle the various tricky policy problems in education (involving difficulties of moderating value decisions, or the difficulties of predicting and evaluating outcomes).

The discussion suggests that the use and production of knowledge could help coming to terms with complexity in education governance. Promoting professionalisation in the realm of teaching could lessen the tension on the system level by managing part of this complexity directly on the level of professional practice.

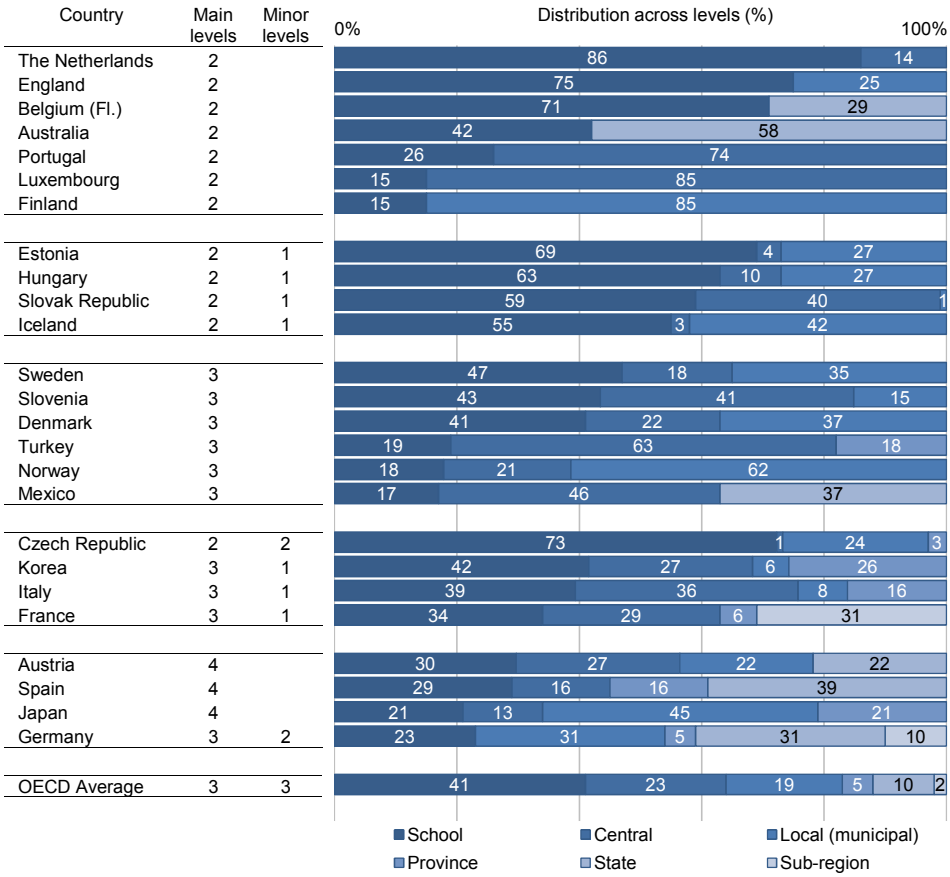
Sources of complexity in a "hybridly" centralised system

The concept of complexity denotes that certain structures might produce unpredictable results and shifts the focus of analysis from uncovering a mechanistic technological machine logic (e.g. a formal bureaucracy, or a "pure" market model) to the understanding of broader and more diverse interrelationships between the involved elements (some of these interrelationships might be notoriously neglected in a mechanistic perspective). A basic assumption of this chapter is that different types of sources of increasing complexity in educational governance and policy making can be reasonably distinguished: (i) Complexity might arise from the various forms of decentralisation, that bring about an increased number and variety of involved actors, and is empirically related also to a strengthening of the stakes of those actors (parents and citizens are more educated and have more self-confidence; diversity in society brings about more diverse interest orientation and less orientation to a common good, and is combined with more diverse interests, etc.). This type of complexity resulting from current changes has been mainly focused in the GCES project so far; it can be called *procedural complexity*. (ii) Another source of complexity might be found in centralised systems, which are formally and legally more or less clearly structured, however, might in practice involve "hybrid" interrelations of different elements of governance, which produce *structural complexities*. This second type is elaborated in this paper.

Throughout the GCES project the dichotomy of centralised-decentralised systems, and the possibilities of measuring it, has often been questioned. One established version of measuring decentralisation has been to look at proportions of decision making at different levels of governance, with the proportion at the school level in decision making in four domains (instruction, personnel, planning/structures, and resources) being taken as an indicator for decentralisation (OECD, 2007). Based on a dichotomous concept of centralisation-decentralisation, an implicit assumption seems often to prevail that decision making at the central level is the main or only complement to the school level. However, the structures are not that simple.

Figure 6.1 gives an overview of governance types based on different distribution of decision making across levels in 26 countries in 2011 (OECD, 2012). From counting and weighting the levels involved, five types can be constructed with different compositions of centralisation-decentralisation, which show that there is no one-dimensional axis. The number of administrative levels involved in decision making varies from two to four levels; the number of levels is multiplying the interrelations between institutions and thus potentially increasing the structural complexity of the governance system.

Figure 6.1. Governance typology, number of decision levels per country



Note: Grouping is based on the number of major and minor levels. The cutting point between main and minor levels has been set at 15% of decision making responsibility; there might be some conceptual overlaps or unclear distinctions between local and sub-regional levels as well as between state and province levels.

Source: Table 6.A1.1 (Annex 6.A1), based on *Education at a Glance 2012* (OECD, 2012).

In Austria, decision making is distributed among four levels (which is typical for much bigger countries, such as Spain, Japan and Germany) and the proportion of decisions at the school level are below average. The responsibilities allocated to the school level are comparatively concentrated on instruction, with weak responsibilities for planning, resources and personnel at this level. Planning is concentrated at the central level and resources and personnel are distributed among the central and the local administrative levels (for further details see Lassnigg and Vogtenhuber, forthcoming). The indicators point to a quite even distribution of responsibilities among four levels (central, regional, local and school); however, the distribution is different for different sectors of schooling:

- Pre-primary education (*Kindergarten* in Austria) is mainly governed at the local level, and also the investment and maintenance of primary schools.
- The lower secondary common track (called *Hauptschule*, and currently being changed towards *Neue Mittelschule*) is more strongly governed at the regional/state (*Länder*) level.
- The academic schools that span compulsory lower and post-compulsory upper secondary education are centrally governed, with administrative responsibilities at the *Länder* level also.
- Post-compulsory fulltime vocational schools (which are strong in Austria beneath apprenticeship)1 are equally governed predominantly at the central level with some administrative responsibilities located at the *Länder* level.

This structure creates gaps at all the main transition points in education, as the authorities also change at these points, making co-ordination more difficult. A main issue of the distribution of responsibilities between different levels are two different categories of teachers, those in compulsory schooling governed by the *Länder* (*Landeslehrer*), and those in post-compulsory schooling governed by the central level (*Bundeslehrer*), comprising different structures of industrial relations, wages, employment conditions, etc. Within compulsory schooling, the primary schools are very widely dispersed across the communes, with many very small communes being responsible for many very small schools.[2]

Table 6.1. Levels of education offered by municipalities, average school and class sizes

Education levels offered	Share of municipalities
Primary only	51%
Primary + lower secondary (common track)	23%
Mixed school structure (types from primary to upper secondary)[1]	15%
No school	12%
Level of education	**School size (students), average**
Primary	107
Lower secondary (common track)	147
Lower secondary (academic track)	402
Upper secondary post-compulsory academic track	264
Upper secondary post-compulsory vocational schools (full-time)	266
Class size in primary education	**Students per class, average**
Overall	18.4
Länder (except Vienna)	17.7
Länder capitals/biggest cities	19.3
Vienna	21.8
Small schools (< 4 classes)	*16.4*

1. Municipalities with at least one post-compulsory school; 6% of communes provide only one post-compulsory school, the remaining 9% provide more than one, the latter include those with more differentiated school structures (overall 203 municipalities).

Source: Author's calculations based on STATISTIK AUSTRIA.

In a trend of consolidation due to budgetary constraints, schools increasingly concentrate in fewer municipalities. Today, about half of all municipalities offer primary education only; further 23% have a primary and lower secondary common school, while more mixed provision of school types, including upper secondary post-compulsory schools are concentrated to only 15% of municipalities. This is reflected in a greater school size at higher levels of education. While the average size of primary schools is about 110 students, academic lower secondary schools average at about 400 students (Table 6.1).

Among primary schools, the dispersed structure leads to a proportion of about 8-10% of all Austrian pupils educated in very small schools comprising less than four classes. Because of different topographical and settlement structures (e.g. degree of urbanisation), as well as different regional policies, the proportion of these small schools varies between *Länder* between 6% and 26%. In very small schools, the class size is about 10% lower than in average-sized schools. About one third of all municipalities (and two thirds of small communes in the countryside, respectively) comprise small schools. They are struggling to a high degree with demographic decline and holding up their school against rationalisation measures.

Interlocking responsibilities and competing governance mechanisms

Two further dimensions of structural complexity must be added from a qualitative perspective. The first is the mode of how the responsibilities are formally allocated to the different governance levels; the second concerns the overall hybridity of the governance system due to its different governance mechanisms.

Firstly, in Austria, the responsibilities regarding education are allocated in a way that they interlock, without clear division of labour between the different levels. The central as well as the regional/state level (*Länder*) both have some legislative and regulatory responsibilities. At the regional level there are two kinds of authorities with interlocking responsibilities: a federal agency, *Landeschulrat*, which is linked to regional politics, and an office of the regional government responsible for schools, *Amt der Landesregierung*. This means that the legal responsibilities are distributed in a complex way so that different governance structures arise in different regions despite their small scale, influenced by the varying political majorities.

Secondly, Austria has another source of hybridity that is even more important and more difficult to grasp than the interlocking responsibilities. It concerns the overlapping of three different types of governance structures, which are differently distributed to different parts of education. The three types are:

- A classical state bureaucracy.
- A federal structure of the nine regions (Länder), comprising individual regional parliaments and governments, which mainly distribute the federal funds in the regional domains.
- A strong system of corporatism, based on interest organisations with to some part compulsory membership (chambers of commerce, chambers of labour, chambers of agriculture).

The governance system combines a quite traditional bureaucratic structure with a kind of distributional federalism that is focused on the distribution of nationally raised tax money to the regional units (*Länder*). The latter have strong democratic political structures (government, parliaments) but very little own money, as the main part of taxes

is raised by the central government. The funds are then distributed via different channels to the regional units (9 *Länder*, about 8 million inhabitants). The distribution is partly based on legal and statutory responsibilities. To an increasing part, the central funds are also distributed via a negotiation process among the units (*Finanzausgleich*), with the money flowing through this mechanism being mostly not earmarked. In addition, Austria has a very high number of municipalities (around 2300), which also are organised with own elected political structures (mayor, parliament) and get their money mainly from the upper political layers (here are the *Länder* an important source). In addition, about 80-90 districts exist, which are a nationally governed administrative structure, themselves not being democratically organised.

As a key responsibility, the *Länder* bodies select and employ the teachers in compulsory schools. In several aspects of education, the *Länder* also have to create supplementary laws for implementation (*Ausführungsgesetzgebung*). Control and inspection structures are scattered on the different levels, and they do not have clear targets to enforce by their work. The schools work mainly as dependent administrative units (*nachgeordnete Dienststellen*) in this bureaucratic-federalist system and are highly regulated. In addition, the public service law, under which the teachers work, is very complex: teachers' work relations are negotiated between the authorities and a differentiated structure of trade unions (regionally, and by school types); for example, in the negotiations of a new law the ministry representatives discuss with more than 20 different trade unions' representatives.

Interrelation of the bureaucracy and federalism

The "hybridity" in educational governance lies, firstly, in the interrelation of the bureaucracy and federalism, which means that the overall bureaucratic structure does not reach from the central level to the schools, but is broken at the regional level, as the central decisions have to be modified and implemented at the *Länder* level. The administrative structure of two parallel regional bodies for school administration, a federal one and one situated at the regional government, creates in fact many political overlaps, so criss-crossing loyalties arise at the *Länder* level that make it difficult that central decisions can reach the delivery level at schools. A main example of this broken structure is the management of the teachers in compulsory school. The *Länder* employ the teachers and pay their salaries in advance, whereas the central level has to refund the money from the central taxes. This is a permanent bone of contention, as the central level wants to curb the costs and has given some basic rules of how to allocate the money basically based on estimates of expected student numbers. However, the *Länder* act on their own interests, organising the schools and employing teachers without providing detailed information about their policies. So the bill from the *Länder* to the federal level is always higher than expected, and the federal budget has (reluctantly) to pay.

As Austrian federalism is highly politicised on the one hand, and does not have own financial resources but has to receive (and thus fight for) resources from the federal taxes on the other, the *Länder* have also strong incentives to make different politics for the purpose of "making a difference" and to "serve the regional identities" vis-à-vis the regional electorate (Lassnigg and Vogtenhuber, forthcoming). This interrelationship of political interests and administrative purposes is creating strong contradictions and tensions in the overall structure. The schools are basically embedded in a tight bureaucratic structure, creating quite tight rules for their everyday practices, which have been heavily documented and criticised since at least the 1990s (Posch and Altrichter,

1993; Schratz and Hartmann, 2009). Although the formal bureaucratic rules apply, they cannot even play out their (potential) strengths, because the political interests of federalism interfere, and the overall result cannot be controlled by the bureaucracy. This structure can be seen as the transformation of a central bureaucracy into nine regional bureaucracies, which undermine the overall coherence of policies. A recent example has been the structural reform of the lower secondary compulsory common school that should change the tracked structure of achievement levels in the *Hauptschule* (HS) into a more integrative structure based on individualisation in *Neue Mittelschule* (NMS). Based on different political majorities with different ideologies towards tracking and differentiation, several different strategies of implementation have emerged at the *Länder* level, which led to the result that the main ideas of NMS were only implemented in a quite small minority of schools (Eder et al., 2015). That is, in the prevailing structure the actors suffer from the negative aspects of the bureaucracy (little discretion in many things, and tight rules and long reaction periods), while its potential strengths of a rational and coherent policy cannot be realised.

As has been indicated above, this kind of structure is differently applied to the different sectors of education: post-compulsory school education is governed by the central level, whereas compulsory education and primary education are under mixed responsibilities, with a different governance structure in each. As the later cycles depend on the earlier ones, the "broken" bureaucratic structure in compulsory education influences also the more centralised post-compulsory system. As an example, the compulsory schools provide the "pipeline" for the transition into post-compulsory education, so the prevailing different regional structures, which reflect rather political preferences than regional conditions, set the frame for the next stage of education, and thus also influence the opportunities of the next generation. In terms of subsidiarity, that is, the idea that things should be managed as near as possible to the practice level, the *Länder* level seems not the most feasible one. For example, the four (relatively) large regions Vienna, Lower and Upper Austria, and Styria have to some extent contrasting or conflicting interests (in particular Vienna as a strongly growing region would need much more resources which are difficult to obtain in the negotiation processes). Additionally, these three non-metropolitan regions are very diverse in themselves, comprising strongly urbanised parts as well as rural areas. These diversities are shaded behind the overall interests of the *Länder*, and policy issues across the *Länder* are not really addressed in this structure (until recently cross-regional exchange of information was very scarce; now Statistics Austria has improved accessible statistics at the levels of communes and also of urban regions).

Corporatist structure in post-compulsory vocational education

The corporatist structure provides a second dimension of the hybridity by its high influence on the apprenticeship part of upper secondary education. Vocational education and training (VET) is dualistic in Austria in the sense that a centralised and bureaucratically governed full-time school system exists in parallel with a classical strongly decentral enterprise-based apprenticeship system that also includes a compulsory part-time school for apprentices (Lassnigg, 2011). Thus at the end of compulsory school, two different systems of about equal size exist, which are differently and separately governed, and in times of demographic change compete for young people. The organisational structure of the corporatist governance is also strongly related to the federalist system with the regional chambers of commerce holding the main administrative responsibilities in the apprenticeship system. As a result of the complex

working of the "collective skills system" in apprenticeship (Busemeyer and Trampusch, 2011), very different political relationships arise in different regions, based on industry structures and cooperative orientations of the actors.

Contrasting this "structural complexity" in a fairly centralised system with the "procedural complexity" in a decentralised system, questions about the different degrees of complexity can be asked. In the structurally complex system comprising the different layers of governance, bureaucracy, federalism and corporatism, the question might be posed, how these different layers are coordinated.[3] In the small and traditionally centralised country, the same actors are contributing to the different overlapping governance mechanisms, which constitute complex varieties of actor constellations. The question might be posed: to which degree an overall coordinated "governability" is possible in this system, where the same actors might develop different positions and orientations according to how they actually act: as a bureaucrat (or official), as a regional policy maker defending the powers of federalism, or as a player in a corporatist interest organisation. Moreover, it must be considered that the driving forces of the "procedural complexity" are also in play in the structurally complex system, as more stakeholders try to bring their stronger stakes into this environment also. Several initiatives to mobilise the civil society towards educational reform can be observed during recent decades (e.g. a referendum in November 2011, or a new initiative by the Federation of Austrian Industries).[4] Each of the three types of governance (bureaucracy, federalism, and corporatism) has been heavily contested for decades; however, because of the multiple and interlocking interests and the many existing veto points change is quite impossible to achieve.

Policy and politics, "hard" and "soft" policy making

This section relates the basic centralised and hybrid governance structure to certain patterns and dynamics of policy making. The concept of governance refers to a wider structural framework and includes a range of actors additionally to specialised policy makers. Policy making in its narrow sense takes place within the governance structure. Consequently, policy-making is directly influenced by a number of other actors, which are to a large number the objects of policy making. The governance structures are a part of the polity, but additionally include various other practices – in our case the education practices. While these practices are not genuinely political, they interact with and influence the policy making process. Embedded into the governance structure, policy making hence does not only follow its own reasoning (that is, designing and implementing adequate, efficient and effective policies) but is subject to political processes with their very own logics.

The distinction between policy and politics has been very much related to the emergence of policy research, which has attempted to analyse with various purposes – from understanding to advise advocacy – the content and conditions of the provisions of political interventions towards the various functions and sectors of society, often called policy fields. In this distinction, education as a practice field can be devised as a policy field that contains its specific topics and challenges. Policy analysis contributes to the field specific understanding and to proposals for solutions, however, this concerns only part of policy making, as the main political decisions concerning a genuine policy field are taken outside of it at a genuine political and government level, and (have to) consider much wider issues and rationalities. As a result, proposals that look very promising at the level of a policy field are not taken over and decided at the level of politics. The

interrelation of the two elements is thus a very tricky and contested issue that in the one or other way contributes to complexity. Concerning the question of this paper about the different kinds of complexity in centralised vs. decentralised governance systems, the relationship between policy and politics is considered an important issue, as these different kinds of regimes might be differently related to the aspects of policy and politics, with centralised governance structures being more strongly tied to politics, giving less leeway to policy proposals.

The distinction between politics and policy has gradually emerged in political science, mostly in combination with the development of various approaches and techniques of empirical policy analysis, evaluation and monitoring. Which policies are appropriate, which are most effective or most efficient to reach certain goals? What is the meaning of certain goals in certain contexts? When these kinds of questions were asked and answered in policy analysis, certain constraints in the overall field of policy making became increasingly clear: the best "rational" answers or solutions from policy research in a certain field conflicted with other dimensions of policy making, and at this edge the distinctions between policy on the one hand, and politics and the polity on the other become important.

The democratic polity conflicts with the inclination of technocratic policy advice based on evidence, and moreover, the constraints in politics of acquiring power or to stay in power are setting the context for field specific policy solutions. Renate Mayntz (2009: 5, Engl. abstract) has posed the contradictions between policy and politics by questioning the possibilities of good policy advice in facilitating "the making of evidence-based and effective policy decisions". Even the best advice and also sound "guidelines for the behaviour of those seeking advice could not assure that advice is used as intended. The effectiveness of policy advice is compromised by the inseparability of Policy and Politics." This inseparability realises when a policy proposal should be implemented, then power as a source of politics comes into play, first as a source for implementation, and second as an aspect of the self-interest of politicians. At this point a policy can be reversed into its opposite, in German Mayntz puts it drastically: "politisch brauchbar kann aber gerade das sein, was wissenschaftlich unhaltbar ist" [what is politically usable could exactly be what is scientifically untenable] (Mayntz, 2009: 13).

The well-established distinction between power oriented *politics* and more technically topic-oriented *policy* (Treib, Bähr and Falkner, 2005; May and Jochim, 2013) seems particularly linked to different governance regimes at the centralisation-decentralisation continuum. This question concerns the relationship between policy and politics, because the structures of governance can only be changed by politics, and consequently, if these changes are necessary, politics must be set in motion. On the other hand, complexity is involved, because politics follow different logics than policies (e.g. creating voters' acceptance or demonstrating competency in order to stay in power vs. good technical solutions to practical problems).

Different approaches in politics vs. policy towards the structural complexity of education governance in Austria might serve as an example of this distinction. At the political level the issue of formal regulatory simplification is mainly addressed, with different powers (regional vs. federal authorities) trying to shift the responsibility towards their own realm without proofing their stance according to efficacy or efficiency. As centralism is currently outmoded, arguments towards decentralisation are strongly emphasised without much resistance. At the policy level, main assertions are problems of keeping a coherent and accountable system vis-à-vis a centrifugal federalism, and

questions of how to develop a framework of accountable school autonomy in a completely federalist system driven by the interest of contrasting each other. The main technical question arises, whether the given practices in the bureaucratic-federalist system will allow for the necessary coherence of education in a small state if the responsibilities would be shifted mainly to the regions.[5]

"Hard" and "soft" policy making

A key point in the GCES-project concerns the strategic potential of the centre to develop and implement coherent policies towards commonly agreed goals in a decentralised governance system. Given the deemed importance of education to further broader national goals of competitiveness and social purposes, many observers are worried that the process of decentralisation might endanger possibilities to further these purposes. Some have posed the question of whether the trend towards decentralisation might have gone too far already. Thus, a main question of the project is how decentralisation should be complemented at the national level to allow for a successful national education policy. The shift from "hard" to "soft" policy making is to some extent seen as a solution for these tasks or problems.

The European "Open mode of coordination (OMC)" serves as an elaborate model of "soft" policy making. Instead of "hard" legal instruments a mechanism of setting goals, evaluating and comparing results according to these goals based on indicators, and providing peer learning has been developed, which should work through influencing, and "naming and shaming" through various kinds of reporting procedures (see Wilkoszewski and Sundby, 2014, for a fuller discussion).The basic setting of the OMC is that the member states have agreed on a set of quantified goals that should be reached at a point in time, and the centre has not the authority to enforce implementation by prescription or sanctions.

The situation is similar in a federalist system, where the responsibility lies at the regional level, and the central authority wants to guide the regional authorities towards certain goals. In Austria, the European policies of the Social Fund (ESF) or the Employment Strategy have provided models for this kind of policymaking, and more recently these kinds of policies have been taken over at the national level. In education some recent examples include the reform of the lower secondary school towards more comprehensiveness (*Neue Mittelschule)*: in this reform substantial additional resources, namely a second teacher for team-teaching or support of individual students in "achievement subjects" were provided for schools that opted into the new structure. To individualise teaching towards the different needs, the common school (*Hauptschule*) should change its instruction methods from institutional differentiation by three achievement levels in main subjects towards instruction in heterogeneous groups without formal differentiation. Within a few years, this change was widely implemented, however, without controlling for instructional and achievement changes; only the input-sided institutional change was observed, and a substantial increase of resources was provided without looking at the results of these changes.

The reform of the lower secondary education towards more comprehensiveness (*Neue Mittelschule*) was implemented in the field of compulsory schooling, where interlocking responsibilities between the central and the regional level prevail. In this context, the responsibility of evaluation is not clear and must be negotiated in the hybrid system. The central level as catalyst and provider of resources would have to involve all the other players in such an activity – however, as the purpose of the reform towards the

establishment of a more comprehensive structure was (and still is) politically disputed among the actors at the different levels, strict obligations for evaluation could have hindered the acceptance and implementation of the policy (it does not, however, hinder the reluctant actors now to criticise the lack of evaluation and question the mainstreaming of *Neue Mittelschule*). Additionally, the reform tried to bridge the different governance levels by demanding a co-operation of teachers employed by the regional authorities with teachers employed by the central authorities (the programme was only implemented in the track of the common "mass" school *Hauptschule*, the academic "elite" track of *Allgemeinbildende Höhere Schule* did not participate in the programme).[6]

Soft policy making in centralised systems and hard policy making in decentralised systems

Concerning the relationship of "hard" and "soft" policy-making on the one hand, and the centralisation-decentralisation dichotomy on the other, Austria provides examples that "soft" policy-making might be used to handle problems in a structurally complex (fairly) centralised system. On this background, the relationship between the following dimensions can be explored (Table 6.2).

Table 6.2. Schematic representation of systems and policy making types

	Centralised system	Decentralised system
Hard policy making	***Bureaucracy*** Prescription and control of activities Standardised resources (qualifications) Mandatory provision	Setting of mandatory goals, Obligatory control of results
Soft policy making	Mobilising commitment, engagement	***Market*** Self-organisation, autonomous activities Flexible resources, acquisition Intervention through incentives, sanctions

The schematic account in Table 6.2 shows on the one hand the "streamlined" relationship of hard policy making in centralised systems (summarised by the bureaucracy) and of soft policy making in decentralised systems (summarised by the market) on the other. More interesting seem the remaining alternatives, soft policy making in centralised systems and hard policy making in decentralised systems. On the background of the Austrian examples, different kinds of alternatives can be devised from this table, and a much wider range of alternative strategic paths can be devised in addition to the main discourses about the centralisation-decentralisation alternative. Relating to proposals from the literature, a stylised elaboration of these alternatives, and some speculation about their implications can be given. In addition to a widely proposed and debated shift from a centralised system to a decentralised system, another alternative – inspired by the above examples from Austria – can be seen in the development of soft policy making as an amendment in centralised systems.

Starting change from decentralised systems, the stylised alternatives are either a shift to centralisation (which is not very much taken into account currently), or the establishment of hard policies in decentralised systems, which is the standard proposal from institutional economics since some time (e.g. Bishop and Wößmann, 2004), and serves as a kind of mainstream path of governance reform. The prototypical policy in this

path of hard policies in decentralised education systems is the use of "high stakes", i.e. complementing decentralised provision at the practice level by procedures of assessment that have clear consequences for the actors.

An interesting question at this stage of reasoning is how the many soft elements towards professionalisation in reform proposals might fit into this scheme. Communication and trust, as well as information, dialogue and capacity building are mentioned as key soft factors in the improvement of governance (cf. Fullan, 2011).[7] Accountability as a key ingredient is closely related to information and communication, and strategic thinking needs capacity building as a key ingredient. These elements are clearly needed in soft policy making, whereas hard policy making has the tendency to make itself immune or invulnerable from these soft elements (e.g. by high stakes policies, which should shift the incentive structures to which the actors at the practice level should react automatically), and thus does not have a high priority to strengthen them more than to an absolutely necessary minimum. Some implications concerning the soft policy elements in the non-mainstream policy paths can be devised as follows:

- Hard policy making in a decentralised environment needs good information and communication about results (accountability), and a high degree of strategic thinking for an appropriate use of incentives and/or sanctions. The overall shift towards hard policy making in soft decentralised systems might undermine trust by the increase of control, building up new instruments and mechanisms for this purpose. It is well known since some time that the governance reforms at the university level point much into this direction, by the building up of new managerial systems and personnel (de Boer, Enders and Schimank, 2007).
- Soft policy making as a complement to hard policy making in a centralised environment depends more strongly on the mentioned soft elements. The question here would be, to which extent an environment of hard policy making provides the necessary and sufficient conditions for the development of the soft ingredients, as they are information, evidence, communication-dialogue, capacity building, and trust. This question refers very much to the issues of knowledge production in centralised systems (e.g. a structural tendency towards the control of knowledge production and flows by the political level might be expected, a tendency which is strongly prevalent in Austria, but might be less so in other centralised systems).

Based on the understanding of the political processes in the actor network embedded in the complex Austrian system, some further questions about how "structures compartmentalize issues" (Burns, 2013: 7), and how the relationship between the structures on the one hand and the soft factors of dialogue, evidence, capacity building, etc. on the other hand might be understood. In more activity related policy proposals geared towards decentralisation it is often stated (also in the process of the GCES project, see Chapter 1), that structures might be important, but were less important than the other elements. The question would be, whether and to which degree structures might systematically condition the other dimensions. Put very bluntly, structures that impose a high degree of centralisation, regulation and (nominal) control are geared towards politics, and are open neither for dialogue, nor for evidence nor for capacity building. Dialogue is restricted by the strong politicisation, evidence is not necessary because the procedures and authorities are clear, and capacity building is restricted to what is prescribed – in effect the soft factors must be somehow processed against the structures.

Change of governance structures in relation to change of governance practices might be particularly tricky, if there are tight formal structures in place (bureaucracy and federalism, which includes a high degree of politicisation). In these structures, a kind of "double bind" arises, as the structures are formally tight, however, practices differ more or less from the formal structures and change has somehow to be imposed in this gap between formal structures and informal practices. For example, a tightly controlled environment can prohibit the large-scale implementation of needed changes. To introduce professionally adequate changes nevertheless, policies for change might be rolled out repeatedly under the label of policy trials at the school level (*Schulversuche*).

So the question might be asked how much energy is absorbed at the various levels by this kind of "double bind" between the obligation as a civil servant to follow the law/rules and the obligation as a professional to achieve substantial results which might be inhibited by the rules. The term is inspired by the "double bind" as a communicational structure, which has been theorized as a source of serious mental disorder decades ago (see Gibney, 2006). Another aspect concerns the well-known phenomenon in education of "too much innovation and too little scaling up", which might in fact also be caused by too tight structures; however, (too) loose structures might also indirectly inhibit innovation because it might not be visible.

Summarising these thoughts, "the shadow of hierarchy" (Peters, 2011: 7), and its consequences for governance would deserve more attention, in particular in relation to the problems of the "disempowerment" of the state. Much energy of reform discourses might be bound in these tensions, and in case of a lack of formal organisational alternatives, the debate tends to be trapped in the politicised state vs. market discourses.

Dialogue, and the issue of a change of mind-sets, and the necessity of creating an infrastructure for this was strongly emphasised in the course of the GCES project, and attributed to capacity building or to governance. These issues reflect the whole topic of agenda setting in policies and politics and of creating political objects, to which a substantive literature exists (e.g. March and Olsen, 1995). Especially for politics, this process is critical, as the "created policy objects" are key for how success and failure is estimated in the public. So a very high interest to control the discourses by politics must exist, and this seems to be related to the structure of the governance system.

Based on this reasoning, we can derive the hypothesis that the more the structure is centralized and politicised the higher the inclination to control the discourses would be, and to this situation the public / the audience might react by not taking the dialogue as a serious one, and to react strategically. Therefore, a situation arises where the dialogue seen as a main instrument for creating trust is foreclosed by the structural conditions driven by distrust and endemic conflict between fundamental positions. How to escape from these self-reinforcing cycles is a challenge for multilevel governance. To disclose this situation and the communicational traps included by detailed discourse analysis could be helpful activities in this situation (as in the case of the "double bind", an element is that the contradictory setting must be negated by the actors that the mechanism works).

Concluding remarks and outlook

The paper has explored sources of complexity in a centralised system, taking Austria as a case. First "structural complexity" in a centralised system was confronted with "procedural complexity" in decentralised systems. As a result, it was argued that in centralised systems the sources of procedural complexity are in place as well and the

sources of structural complexity exist in addition to that. Structural issues concern politics, which per definition overrule policy.

Politics is in play in different ways, first centralised systems create the notion (or illusion) that the governance structure is a machine-like transformation mechanism that brings the political decisions more or less straightforwardly to the ground, so politics must take the right decisions and fuel them into the "machine" – it was demonstrated that centralised structures might include sources of complexity that counter this straightforward notion and might block the system. In the Austrian case this is the self-binding of politics to a high degree of consensus and the interrelation between the bureaucracy and federalism. If problems of this kind are detected, the second role of politics comes into play, which states that the change of structures requires political decisions. As long as decisions towards the change of structures are not taken, policy proposals and practices can only work within the given structures. At the same time, politics is to some degree likewise trapped in structural complexities, as the exploration of the Austrian experience shows (even if federalism is considered as highly inefficient, it is there and, in its extreme, if it should be abolished, its stakeholders must abolish themselves, what clearly is an unrealistic demand, easily to be seen on the occasion of the fierce resistance against mergers between small communes).

Contrasting approaches, based on analyses of governance in the United States, are theorising the reverse direction of channels of influence by feedback from policy to politics (May and Jochim, 2013). This direction is not analysed in this paper, however, these feedback processes can be expected to work differently at the centralised end of the continuum than at the decentralised one: In a decentralised system, a degree of diversity of solutions is welcome and more or less "part of the game". Here, different solutions can compete and some degree of evolutionary change towards successful solutions is expected. In a centralised system, prescriptions work towards conformity and diversity works against the rules. To which degree emergent processes based on diversity are working towards change also in highly regulated or centralised systems is a big issue in research, particularly in historical institutionalism (Mahoney and Thelen, 2010). In the Austrian development of education, the necessary room for initiatives to manoeuvre within the established structures, and consequently their potential for a change of practice, are a longstanding topic of debate. Ideas to increase this room to move have been supported time and again – however, these attempts seem not to have been successful so far.

The second section looked at the relationship between hard and soft policy making and the centralised-decentralised dichotomy. A simple cross-tabulation of these dimensions guided the attention from the main diagonal of the table to the secondary diagonal of soft policy making in centralised systems and hard policy making in decentralised systems. It was shown that a main current trend can be seen in the development of hard policy making in decentralised systems, e.g. by control of results through "high stakes" policies. From hard policy making in centralised systems three different policy alternatives can be seen in this framework: (i) a shift to a decentralised system, (ii) a doubling of hard policy making by adding the control of results to the control of the inputs and procedures, and (iii) by adding soft policy making to hard policy making in the centralised system. The third alternative seems particularly interesting.

At first sight there seems to be a marriage between centralised systems and hard governance on the one hand and between decentralised systems and soft governance on the other – a shift from one governance regime to the other, if it can be made, would also

change the practices. The closer look has shown that things are not that easy, as there are criss-crossing relationships and also to some extent mixed practices. What these interrelations mean in terms of complexity is not easy to answer. It will clearly increase complicacy, as substantial new elements are added to the existing practices (e.g. the testing procedures and their utilisation channels with hard policy in decentralised systems, or the necessary consultation and observation mechanisms if soft policy is added to hard politics). Whether this would also increase complexity in terms of unpredictability is an open question so far. Concerning the establishment of high stakes there are conflicting results available. Reforming the production and use of knowledge towards professionalisation of education practice could alleviate some issues pertaining to complexity, by integrating complexity into the professional practice.

Reforming production and use of knowledge to cope with complexity

The production and use of knowledge seems to be a key element in the analysed interrelations which deserves more analysis and attention. The argument was started with the different epistemological approaches of considering complicacy and complexity with the focus on the issue of (un)predictability. In principle, knowledge production should increase the predictability of how the system works, and the conceptual shift towards outcome orientation has somehow refocused the knowledge production in education towards a more functional view of how to assess and improve the outcomes.

Large scale assessments have brought a new emphasis on knowledge production that is situated externally to the actors in the system, and with the providers of this knowledge the assessments have added also a new class of actors into the system – the analysts and researchers/developers – which in some respect knows more than the actors involved about their work (know more must not in any case mean know better). Nevertheless the actors own their knowledge as previously, and a main question concerns how the new category of knowledge – and its providers respectively – is processed and integrated. Here a basic contention is that the policy makers own and need different knowledge than the educational practitioners, and that this constitutes different relationships between those actors' categories and the researchers and analysts, which might lead to confusion if not distinguished appropriately.

Obviously there are disagreements about the new knowledge practices, e.g. to which extent the assessments actually represent what a system achieves, or how the new knowledge might or should be used in the practices of the actors. A main issue of predictability concerns the "production function", i.e. what is known about how the outcomes can be improved, and which actions might trigger improvement. The existing proposals for improvement are not trivial and can in turn be questioned towards the topic of complexity and predictability, as they reproduce overall accounts of the complexity of the systems in some respect: In a simplified manner we can distinguish, firstly, economic proposals, which focus on the distribution of incentives among the actors; secondly, managerial proposals, which focus on procedures of quality assurance and improvement; and thirdly, genuine pedagogical proposals, which focus on teachers and their competences.

In the argument of this paper we can say that complex proposals meet complex systems, and what will come out of this is notoriously unpredictable. In the discourses some say it is the teachers, others say it is the incentive structure; again others say it is the management, and in fact it can be expected that all factors contribute somehow. This "somehow" is the point where governance comes into play, as the governance system

combines the actors in a certain way, and it determines to some extent how the decision making about alternatives works in a system. Relating this reasoning to the above argument and analysis, this poses the question whether the different governance structures create different channels of how knowledge flows and can be used in a system for ongoing practices as well as for a change of practices, if deemed necessary.

Based on the Austrian experience, the knowledge production and flow appears to be controlled to a higher degree by politics in centralised governance systems than in decentralised systems. That is, while education research can be close to policy advice, it is less likely to find its way into education practice, as the logics of politics tend (e.g. ideology, power politics) to prevail over technical policy making. In decentralised systems the relationships might be more open, with an interest of policy makers to gather knowledge about the more diverse and distributed system. Nevertheless, the flow between research and educational practice might in both systems be weak: It might be more unpredictable in decentralised systems, and more constrained in centralised systems. As argued above, the relationship between policy and politics will be different in centralised and decentralised systems, with a stronger potential of policy in the latter. To improve the impact of external knowledge production, these relationships should be analysed more thoroughly.

Some main approaches in systems theory, in particular based on the work of Niklas Luhmann (1990), rest strongly on the concept of *autopoiesis*. These approaches theorize the political system as a system besides the others, without being privileged to really control the other systems (with similarities to ideas of institutionalism); rather, the political system must try to condition the remaining systems. It is theorized that each system has its own logic (*autopoiesis*) and communicates with other systems via contingent coupling mechanisms. This view has important consequences for the understanding of governance in centralised systems, as it particularly emphasises the gap between policy and politics: policy sits at the intersection of research and politics, but has to follow primarily the logics of politics (see Stichweh, 2011, for an overview, also Mayntz and Scharpf, 2005, problematising this view). Politics on the one hand and education research and policy advice on the other follow the logics of different systems, and will only occasionally strongly act in the same direction ("windows of opportunity"). This approach of a systemic view also brings up the issue of how the knowledge can flow from research to the practitioners and teachers, and consequently, how learning can be facilitated in education practice, as well as how different forms and modes of knowledge (e.g. research or practice generated) can "talk to each other" in this respect.

A more thorough analysis of how forms of knowledge interact, in particular how research knowledge flows and combines with the other forms of knowledge would be necessary in order to understand the potential impact. Secondly, a closer look at the distinct flows of knowledge and their potentials would be helpful, analysing and comparing the type of direct flows of knowledge between research and practice (in both directions), and another type of flows also finally between research and practice, but mediated by policy and politics. To understand the consequences for governance of these two types of knowledge flows could improve the use of knowledge.

Notes

1. The governance of apprenticeship is mostly separate from school governance, run by another Ministry (Economic Affairs) and the Social Partners. Only the compulsory part-time school that apprentices must attend is under the responsibility of the school governance system, also distributed among the federal and the *Länder* level.
2. About 60% of all communes are very small and comprise a population below 2000, and of those three quarters are situated in the countryside (where consequently 70% are below 2000). About 13% of the population lives in those about 1 000 small communes in the countryside (46% of all communes).
3. The issues of co-ordination in federalist systems has been also taken up recently in political science by a set of illuminating case studies; unfortunately Austria is not included in this research so far; see Bolleyer et al., 2014.
4. See www.vbbi.at/; www.iv-net.at/b3487/beste-bildung-fuer-oesterreichs-zukunft-die-inhalte-des-iv-konzepts/ (in German).
5. More recently, this question was radicalised, as a joint proposal of the Ministry and the *Länder* is under way to shift the responsibilities for implementation to the *Länder* level, whereas the institutions at a more local district level should be removed.
6. Policies in other sectors (early education, basic adult education), and the overall "Lifelong Learning – Strategy" also have taken up elements of this kind of "soft policy making".
7. See the presentations at the Paris conference of the GCES project (www.oecd.org/edu/ceri/thirdthematicconferenceannouncement.htm), as well as the material around the approach of the fourth way education reforms, based on Hargreaves and Shirley (2009).

References

Bishop, J.H. and L. Wößmann (2004), "Institutional effects in a simple model of educational production", *Education Economics,* Vol. 12/1, pp. 17-38.

De Boer, H., J. Enders and U. Schimank (2007), "On the way towards new public management? The governance of university Systems in England, the Netherlands, Austria, and Germany", in D. Jansen, (ed.) *New Forms of Governance in Research Organizations: Disciplinary Approaches, Interfaces and Integration*, Springer, Dordrecht, pp. 137-152.

Bolleyer, N., W. Swenden and N. McEwen (2014), "Introduction, A theoretical perspective on multi-level systems in Europe: Constitutional power and partisan conflict", *Comparative European Politics*, Vol. 12/4-5, pp. 367-383.

Burns, T. (2013), "Governing complex education systems", Presentation at 3rd GCES Conference, Effective Multilevel Governance in Education, Paris, 17-18 June 2013.

Busemeyer, M.R. and C. Trampusch (eds.) (2011), *The Political Economy of Collective Skill Formation,* Oxford University Press, Oxford.

Eder, F. et al. (eds.) (2015), "Evaluation der Neuen Mittelschule (NMS) – Befunde aus den Anfangskohorten" [Evaluation of the Neue Mittelschule (NMS) – Findings from the Initial Cohorts], Research Report, Universität Salzburg; Johannes Kepler Universität Linz and Pädagogische HochschuleOberösterreich, Salzburg and Linz, www.bmbf.gv.at/schulen/bw/nms/eval_forschungsbericht.pdf?4safj1.

Fullan, M. (2011), "Choosing the wrong drivers for whole system reform", *Centre for Strategic Education Seminar Series Paper No. 204* (May), East Melbourne, Victoria www.edsource.org/wp-content/uploads/Fullan-Wrong-Drivers1.pdf.

Gibney, P. (2006), "The double bind theory: Still crazy-making after all these years", *Psychotherapy in Australia*, Vol. 12/3, pp. 48-55.

Hargreaves, A. and D. Shirley (2009), *The Fourth Way: The Inspiring Future for Educational Change*, Corwin Press, Thousand Oaks, CA.

Lassnigg, L. (2011), "The 'duality' of VET in Austria: institutional competition between school and apprenticeship", *Journal of Vocational Education and Training*, Vol. 63/3, pp. 417-438.

Lassnigg, L. (2009), "Policy learning and outcome orientation in Austrian VET-Theoretical and empirical explorations", in F.K. Oser et al. (eds), *VET Boost: Towards a Theory of Professional Competencies: Essays in Honor of Frank Achtenhagen*, Sense Publishers, Rotterdam/Chinese Taipei, pp. 385-408.

Lassnigg, L. et al. (2007), "Ökonomische Bewertung der Struktur und Effizienz des österreichischen Bildungswesens und seiner Verwaltung" [Economic assessment of the structure and efficiency of the Austrian school system and its administration], IHS research report commissioned by Federal Ministry of Education, Arts and Culture, www.equi.at/dateien/ihs_oekbew.pdf.

Lassnigg, L. and S. Vogtenhuber (forthcoming), "Challenges in Austrian educational governance revisited", IHS-Sociological Series, No. 107, Institute for Advanced Studies, Vienna, Austria.

Luhmann, N. (1990), *Essays on Self-Reference*, Columbia University Press, New York.

Mahoney, J. and K. Thelen (eds.) (2010), *Explaining Institutional Change: Ambiguity, Agency, and Power*, Cambridge University Press, Cambridge.

March, J.G. and J.P. Olsen (1995), *Democratic Governance*, Free Press, New York.

May P., J. Jochim and E. Ashley (2013), "Policy regime perspectives: Policies, politics, and governing", Paper presented at the International Conference on Public Policy, 26-28 June 2013, Grenoble, France.

Mayntz, R. (2009) "Leitlinien für die Regelung wissenschaftlicher Politikberatung" [*Speaking Truth to Power: Guidelines for the regulation of scientific policy advice*], *dms – der moderne staat – Zeitschrift für Public Policy, Recht und Management*, Vol. 2/1, pp. 5-16.

Mayntz, R. and F.W. Scharpf (2005), "Politische Steuerung – Heute?" [*Political Steering – Today?*] *Zeitschrift für Soziologie*, Vol. 34/ 3, pp. 236-243.

OECD (2012), *Education at a Glance 2012: OECD Indicators*, OECD Publishing, Paris, http://dx.doi.org/10.1787/eag-2012-en.

OECD (2007), "Linkages between performance and institutions in the primary and secondary education sector", ECO/CPE/WP1(2007)4, OECD Economics Department, Economic Policy Committee Working Party No. 1 on Macroeconomic and Structural Policy Analysis, 21 February 2007, www2.oecd.org/oecdinfo/info.aspx?app=OLIScoteEN&Ref=ECO/CPE/WP1(2007)4.

Peters, B.G. (2011), "Steering educational policy: Managing complex and complicated policy processes", Keynote at first GCES-Conference, Effective governance: From the centre, 21-22 November 2011, The Hague, www.oecd.org/edu/ceri/49613505.pdf.

Posch, P. and H. Altrichter (eds.) (1993), "Schulautonomie in Österreich" [School Autonomy in Austria] 2nd revised edition, Bildungsforschung in Österreich [Educational Research in Austria], Vol. 1, Ministry of Education, Vienna, Austria.

Schratz, M. and M. Hartmann (2009), "Schulautonomie in Österreich: Bilanz und Perspektiven für eine eigenverantwortliche Schule" [School Autonomy in Austria: Results and Perspectives for the Autonomous School], Chapter C2 Austrian National Education Report, Vol. 2, www.bifie.at/buch/1024/c/2.

Stichweh, R. and N. Luhmann (2011), *The Wiley-Blackwell Companion to Major Social Theorists: Contemporary Social Theorists*, in G. Ritzer, and J. Stepnisky (eds.), Vol. 2, Wiley-Blackwell, Hoboken, New Jersey.

Treib, O., H. Bähr and G. Falkner (2005), "Modes of governance: A note towards conceptual clarification", *European Governance Papers (EUROGOV)*, No. N-05-02, http://edoc.vifapol.de/opus/volltexte/2011/2458/pdf/egp_newgov_N_05_02.pdf.

Wilkoszewski, H. and E. Sundby (2014), "Steering from the centre: New modes of governance in multi-level education systems", *OECD Education Working Papers*, No. 109, OECD Publishing, Paris, http://dx.doi.org/10.1787/5jxswcfs4s5g-en.

Annex 6.A1

Table 6.A1.1. Governance typology, number of decision levels per country

	Number of levels			Decision levels present in country						Share of decisions per level (%)					
	Main levels	Minor levels	Sum levels	School	Central	Local	Province	State	Sub-region	School	Central	Local	Province	State	Sub-region
Netherlands	2		2	x	x					86	14				
England	2		2	x		x				75		25			
Belgium (Fl.)	2		2	x				x		71				29	
Australia	2		2	x				x		42				58	
Portugal	2		2	x	x					26	74				
Luxembourg	2		2	x	x					15	85				
Finland	2		2	x		x				15		85			
Group 1 (mean)	**2**									**47**	**25**	**16**		**12**	
Estonia	2	1	2+1	x	(x)	x				69	4	27			
Hungary	2	1	2+1	x	(x)	x				63	10	27			
Slovak Republic	2	1	2+1	x	x	(x)				59	40	1			
Iceland	2	1	2+1	x	(x)	x				55	3	42			
Group 2 (mean)	**2**	**1**								**62**	**14**	**24**			
Sweden	3		3	x	x	x				47	18	35			
Slovenia	3		3	x	x	x				43	41	15			
Denmark	3		3	x	x	x				41	22	37			
Turkey	3		3	x	x		x			19	63		18		
Norway	3		3	x	x	x				18	21	62			
Mexico	3		3	x	x			x		17	46			37	
Group 3 (mean)	**3**									**33**	**32**	**27**	**3**	**5**	
Czech	2	2	2+2	x	(x)	x	(x)			73	1	24	3		
Korea	3	1	3+1	x	x	(x)	x			42	27	6	26		
Italy	3	1	3+1	x	x	(x)	x			39	36	8	16		
France	3	1	3+1	x	x		(x)		x	34	29		6		31
Group 4 (mean)	2.75	1.25								47	23	9	13		8

Annex 6.A1 *(cont.)*

Table 6.A1.1. Governance typology, number of decision levels per country

	Number of levels			Decision levels present in country						Share of decisions per level (%)					
	Main levels	Minor levels	Sum levels	School	Central	Local	Province	State	Sub-region	School	Central	Local	Province	State	Sub-region
Austria	4		4	x	x	x		x		30	27	22		22	
Spain	4		4	x	x		x	x		29	16		16	39	
Japan	4		4	x	x	x	x			21	13	45	21		
Germany	3	2	3+2	x		x	(x)	x	(x)	23		31	5	31	10
Group 5 (mean)	3.75	0.5								26	14	24	11	23	2
OECD (mean)	3	3	3+3	x	x	x	(x)	(x)	(x)	41	23	19	5	10	2
Sum				**26**	**17+4**	**14+3**	**5+4**	**6+1**	**1+2**						

Note: Grouping is based on the number of major (indicated by x) and minor levels (indicated by (x) in brackets). The cutting point between main and minor levels has been set at 15% of decision-making responsibility; there might be some conceptual overlaps or unclear distinctions between local and sub-regional levels as well as between state and province levels.

Source: Author's calculations based on *Education at a Glance 2012* (OECD, 2012).

Chapter 7.

Knowledge and research use in local capacity building

Philippa Cordingley

Centre for the Use of Research and Evidence in Education (CUREE)

Knowledge is vital for teacher quality, both in terms of research evidence and practitioner expertise. The chapter describes possible tensions between research knowledge and practitioner knowledge. Issues revolve around practitioners' knowledge lacking distance from the research subject on the one hand and research based knowledge not being usable for practitioners in the busy environment of the school on the other.

Based on a number of examples from England (United Kingdom), the chapter proposes concrete ways to build teacher capacity for engaging with research and to conduct research of their own and increase appreciation of practitioner knowledge in the research community. With regards to education governance, the chapter discusses how policy making can facilitate teachers' motivation and involvement in research by providing the tools for easier use of research knowledge. Importantly, practitioner research should be accompanied by rigorous quality control to ensure fruitful and generalisable findings and provide connecting points with large-scale education research.

Introduction

Key changes affecting use of knowledge and evidence in the context of governance include moves from central regulation to de-regulation, from a small numbers of known actors to an increasingly large and wide-ranging group of stakeholders and changes in authority at different levels, especially related to increases in school autonomy. In the context these changes, this chapter approaches knowledge as implying a status or a warrant for action. In the context of policy making for school improvement it suggests concepts, approaches, phenomena and skills that have been proven to work, usually in a range of contexts and for a significant number of people. It also explores the ways in which teachers approach and understand such warrants and notions of proof through the lens of their engagement with the research of others and in their own research.

The chapter reflects on the kinds of knowledge that supports effective governance of education in two parts. It starts with showing how external research can inform teachers for their own research based on practical classroom knowledge. It uses two examples to contextualise the analysis; the first presenting how advanced teacher-led research can function in practice, while the second illustrates how external research can be made accessible to teachers as a prerequisite for own research. The second part of the chapter locates the analysis of the local level in the in broader structures and processes of governance: How can governance facilitate the knowledge flow between external research and teachers and how can research feed into improved practices? This second part directs the spotlight on multiple issues surrounding this issue both on the supply and the demand side. The discussion and examples from England suggest that capacity building through knowledge and research use among education practitioners is a promising avenue to increase evidence informed professionalism. Most important in this undertaking is to develop intuitive tools enabling and motivating teachers to engage with research in a busy school environment, and adjust the governance structures to ensure opportunities for practitioners to engage with research as well as ensuring the quality of teacher-led research to increase its relevance.

Interactions between knowledge, policy and evidence

The connections between knowledge, policy and evidence are many and disparate. The more complex the system the more complex the linkages will be. But whatever the distribution of decision making and agency, it is increasingly recognised (Mourshed, Chijioke and Barber, 2010) that teacher and teaching quality is the fundamental driver of the quality of student experiences and outcomes. Although the governance and knowledge systems in different countries take many forms, the high level components remain broadly constant.

For example, (the majority of) systems where Initial Teacher Education and qualification is governed at national level and operated through Higher Education Institutions (HEIs), initial teacher education and linked regulation is a key venue for and driver of connecting policy, practice, knowledge and evidence. Increasingly the development of national standards for teachers that specify more advanced standards for professional practice also create a governance platform for increasing the connections between teachers' knowledge, skills and practices and the wider evidence base about effectiveness. Some, such as those developed by the Australian Institute for Teaching and Leadership (AITSL) do so with explicit reference to the development and use of knowledge from research. For example their second standard relating to teachers' understanding about how students learn puts understanding from research as central to

practice in increasingly complex ways at all four levels of progression. So continuing professional development and learning (CPDL) is also becoming an important vector for building capacity by connecting knowledge, evidence and practice.

Knowledge can be either research-based or practice-based. A knowledge based capacity building system needs to acknowledge and work with what teachers know and do and the ways they interpret, fashion and enact professional knowledge. Whilst knowledge from multiple cases, especially when rigorously researched, analysed and effectively summarised has much to offer in terms of quality improvement, it is teachers' knowledge of the pupils, the curriculum and the practical realities of school life that is most on their minds on a day to day basis.

In most education systems, higher education institutions are seen as a key source and purveyor of warranted knowledge, drawing from their own and their colleague's education research. Where teacher educators work in institutions with strong research cultures, or where there are strong connections between teacher educators and research, their mutual influence on each other may be strong and visible. But teachers within higher education institutions, like their colleagues in schools, also draw on other warrants. First, their practices are inevitably and properly also shaped by the systems and policies of their host institutions and the requirements of the governments that fund them. They are also based on the views of recognised thought leaders.

Ideally, the teacher educators' practices are shaped by their connections with the schools where their students will work and within which students' early experiments with practice take place. All this is also heavily influenced by their own professional identities (as, for example, teacher, researcher or teacher educator), by often tacit practice, and above all by colleagues' own internalised beliefs and assumptions and experiences. School teachers' engagement with knowledge and evidence is subject to a similarly wider range of influences, not least the support and structures put in place by school leaders, the needs of, and their aspirations for, their students, the demands made by their parents and colleagues and by those with employment and other regulatory powers over them. How then might the capacity of teachers to engage with knowledge from research in meaningful ways be developed as part of building capacity for improvement in the context of complex governance (see also OECD, 2015)?

It is, of course, this complex network of influences that large-scale research seeks to tease apart in identifying the effectiveness of particular approaches, and the prevalence of particular trends in pupil achievement. To do so, research-based knowledge draws heavily on multiple cases structures for reliability and validity to create something independent, something usable in multiple contexts and therefore context free. Practice-based knowledge addresses the same phenomena but recognises their boundedness to context; it is a more human phenomenon dependent on the people who enact it; metaphorically speaking, whose very act of holding the globe hides some part of it from scrutiny: Research-based knowledge could be considered as a globe representing the earth spinning in space and practice-based knowledge as the same globe being supported by a pair of hands.

If policy makers are to harness research-based knowledge to practice for the purpose of improving schooling and achievement, these two worlds need to be brought together. Both are busy and oriented to their own very different power structures, imperatives, rhythms and realities. Simply telling teachers, school leaders or school boards that X offers an effective approach that is superior to current practices rarely has much impact. For example, Assessment for Learning (AfL) is one of the approaches that has the

strongest international research warrants; yet, whilst widely "known" and practiced, it is still relatively little understood (Black and Wiliam, 1998). As Marshall and Drummond (2006) point out, in the schools in England where the practices had been both extremely popular and assiduously disseminated through National Strategies backed up with extensive Continuing Professional Development (CPD), probably only 20% of the full potential of AfL is actually being deployed by the teachers trying to use it. In particular although teachers are using techniques to learn more about how their pupils are experiencing their lessons; they are not, as yet however, using the information they gain this way to refine the next steps in the learning sequence to build on what they discover. So use of research and harnessing research based knowledge as a tool for improving practice and capacity building is not, as yet, an established art form.

Research-based knowledge as a tool for improving practice and building capacity

This first part of the chapter sets out with an example from England in the United Kingdom, presenting how advanced teacher-led research and its scaling up to the system level can look like in practice (Box 7.1). It discusses the conditions and potential obstacles connected to this approach of teacher-led research and use of knowledge. Prerequisite for engaging into their own research based on evidence from the classroom is teachers' capacity to engage and reflect meaningfully on existing research. Hence, shifting the focus on how to build capacity ultimately enabling teachers to improve teaching practice directly, the section follows with a discussion of the prerequisites for making research accessible and usable for teachers. The section concludes with an illustration (Box 7.2) of how such a capacity building approach would look in practice.

Box 7.1. Advanced teacher-led research and its scaling up in practice

In 1997, the English national Teacher Training Agency (TTA) announced a programme of national research awards for teachers who were willing to carry out enquiries on behalf of the profession. The process would be rigorously quality-assured and, if successful, published in order to both inform teachers of its content and to act as a role model for others teachers in engaging with knowledge developed elsewhere and in their own evidence-informed learning.

One of the successful teachers was Romey Tacon, a head of an infant school in a deprived town in coastal England, who was deeply concerned about the lack of progress of a significant number of pupils in numeracy. Working with a colleague and higher education mentor, Tony Wing from Brighton University, the two teachers explored the findings of Catherine Stern's (Stern, 1949) research into the development of understanding of number relations (Tacon and Wing, 2004). Stern's work focused in particular on visual representation of number relations; the two teachers used this to construct and test apparatus that would support pupil learning, and to develop effective ways of introducing and working with this apparatus. They were delighted by a very positive and swift response from pupils. Quite soon their colleagues began to take an interest, struck by the animated and detailed conversations about changes in the learning of all pupils and of known struggling learners in particular. By the end of the first year of the research word had begun to spread to other local schools and the formal teacher research report published by TTA began to attract wider interest.

A second year grant enabled wider testing of the approach with other neighbouring schools, with other year groups in the host school and in local junior schools. Again the results continued to be impressive. The publication of the findings attracted a good deal of local and regional attention and Romey ran a number of local conferences with support from the Local Authority. When her capacity to support the insatiable demand for places ran out she opened up her school

Box 7.1. Advanced teacher-led research and its scaling up in practice *(cont.)*

on Tuesday afternoons so local teachers and head teachers could come and observe the approach at work and discuss what they had seen with the teachers afterwards. The impact of the approach was particularly striking for struggling learners, especially for those with short term memory challenges, like pupils with Down's syndrome. The Teacher Training Agency put the teachers in touch with a Charitable Foundation able to fund further, larger scale research and they also contacted a leading charity supporting pupils with Down's syndrome. Further regional trials revealed similar striking patterns in improvement and, over time the resources and approach were developed for publication nationally and internationally. The intellectual property behind the approach was eventually sold to a major international publishing house.

The resulting approach, known as Numicon mathematics, is now widely recognised as playing a significant role in advancing the numeracy skills of pupils who struggle with numeracy, especially those with Down's syndrome, in ways that significantly advance the level of progress they can attain and thus the extent to which they can function independently in society. Numicon is now the subject of a large-scale randomised control trial across more than eighty schools in England (Cordingley and Crisp, 2014).

Conditions for teacher-led research and knowledge-related capacity building

The example from England (Box 7.1) describes a number of outcomes linked to a national policy geared to building capacity through research engagement depended on a number of conditions and drivers. The following elements can be distilled from the example described above:

Firstly, it addressed a "wicked issue". A wicked issue refers to a profound and continuing concern to substantial numbers of teachers and the education system as a whole and one that was driven by specific aspirations for pupils. This led to an emphasis on meeting the needs of other potential teacher users of the research from the outset, connecting with teachers' core identity around the business of meeting the learning needs of others. Secondly, there was a strong existing evidence base about effectiveness on which the teacher could build with the help of a mentor who has had a chance to explore the original research findings in a number of contexts. Thirdly, as the work developed incrementally, the teacher was able to tap into a sustained funding stream over a period of three years. Subsequent funding from a separate, charitable organisation enabled the teacher's own research and development to continue and grow to scale until it became self-sustaining with the arrival of a publishing company.

Importantly, the teacher-led research described took place in an environment with compelling responses from pupils, teachers and schools and supported by a policy ensuring funding, recognition and support. In the form of a funding scheme, the policy comprised the following key requirements for teacher-led research:

- The teacher-led research had to build on existing evidence.
- Scientific rigour enabled the policy to fund the research project as a "standard bearer" on behalf of the profession.
- Structures needed to be in place that facilitated systematic engagement with evidence to move the teacher reports beyond the description and opinion that had characterised many teacher case studies up to that point.
- Coaching in methods and in making research accessible to peer and expert scrutiny.

Most but not all of these conditions were designed into the policy to promote research-informed practice of the Teacher Training Agency (TTA) through its Teacher Research Grant Policy. Its success, combined with an increasing recognition of the importance of research-informed practice more generally, acted as a springboard for the creation of a number of other significant funded research programmes for teachers in England.[1] Each of these initiatives built upon a subset of the findings from the first national policy for promoting use of research and evidence. Each programme was addressing a particular subset of issues such as leadership, school-to-school networking, or accreditation of post graduate CPD (Continuing Professional Development). Other differences arose because each initiative was nested within the standard operating protocols and organisational values of the host organisation. Despite such distinctions, an increasingly shared understanding of the key ingredients for promoting evidence-informed professional development to enhance the depth of knowledge and practice can be understood as a recognisable trend from 1997-2009 (Cordingley, 2010).

Helping teachers engage with research

The emergence in England of a mature and increasingly coherent evidence base about what makes a difference to pupils, as well as to teachers in CPD and practitioners' use of research was initiated and supported by Government investment in the development of a methodology for systematic and technical reviewing of research findings (Bell et al., 2010; Cordingley et al., 2007; Cordingley et al., 2005a; Cordingley et al., 2005b; Cordingley et al., 2003; Timperley et al., 2007).

Based on this government-initiated methodology for reviewing education research, a number of English national organisations (the General Teaching Council, the Department for Education, the National Teacher Research Panel, CUREE and the Learning and Skills Improvement Service) cooperated in funding a systematic review of the evidence about the full spectrum of teacher use of research. The review explored the evidence about how teachers engage with existing research and its effect and how this is similar to and different from their engagement in their own research. This review also explored how teacher engagement in and with research compares with the experiences of health and social care professionals.

Taken together, the use of research and CPD reviews (Bell et al., 2010; Cordingley et al., 2007; Cordingley et al., 2005a; Cordingley et al., 2005b; Cordingley et al., 2003; Timperley et al., 2007) encompass almost 50 000 studies which were filtered to identify approximately 250 studies that provide high quality and relevant evidence. From these studies data could be extracted and synthesised to identify the most common problems associated with teacher research, the potential benefits of functioning engagement of teachers in research and the key characteristics needed to help teachers to engage in and with research. The major findings are as follows:

Regarding difficulties experienced by teachers in relation to research, two dimensions were most frequently found: time and inadequate facilitation/external support. With respect to the first, teachers frequently reported a lack of time to familiarize themselves with new strategies and time for interpreting and adapting the approaches to their specific context. Additionally, a lack of time frequently appeared to lead to an overload of information or distraction: teachers struggled to engage in their own research or with external research in sufficient depth.

In terms of inadequate facilitation and/or external support, teachers frequently reported problems related to insufficient support at the point it was needed; too little contact with experts regarding the content of new approaches and learning to use them; as well as a lack of practical ways to structure experimentation and adapting new approaches for specific contexts. Other problems pertained to poor research instruments, for example over-elaborate and lengthy surveys. On the administrative side, teacher reported a shortage of practical help with enquiry processes such as data entry, typing up interviews, coding data, managing videos and surveys.

With respect to positive outcomes, reviews highlighted that the benefits of building capacity through engagement in and with research are significant both for pupils and for teachers. For pupils, the research reviews highlight links between such activities and pupils' motivation, their attitudes to different subjects, test performance and specific skills (e.g. questioning skills). Similarly, the reviews highlight links to pupils' self- and group-organisation such as their approaches to collaboration and the selection of learning/ problem solving strategies.

For teachers there are links with improvements in their self-confidence, for example related to risk-taking and efficacy; teachers' willingness and ability to change practice; improvement in subject and pedagogy knowledge and using these skills to match pupils' needs; as well as teachers' increased willingness to engage in continued professional learning.

The findings across all these reviews (Bell et al., 2010; Cordingley et al., 2007; Cordingley et al., 2005a; Cordingley et al., 2005b; Cordingley et al., 2003; Timperley et al., 2007) are remarkably consistent and highlight a number of key characteristics for evidence-informed practice (Table 7.1).

Table 7.1. Key characteristics needed for evidence informed practice

Key characteristics of evidence informed practice gathered from reviews	
Continuing specialist support	• Training, including instruction in the essential core of new approaches and facilitation of the development of an understanding of the key principles underpinning those approaches. • Modelling demonstrating innovative strategies at work in a range of settings and contexts and practicing what is being preached. • Guidance and critical friendship to challenge orthodoxies and expand views about what is possible on a sustained basis – sometimes called coaching, or mentoring, sometimes collaborative enquiry. • Tools and frameworks such as observations frameworks to support learning from looking, analysis grids and planning tools to secure consistency and coherence.
Continuing peer support	• Professional learners make themselves reciprocally vulnerable thus increasing ownership, commitment and a willingness to take risks. • Peer support speeds up the process of developing trust that enables unlearning of old assumptions and habits as well as the development of new understandings and practices.
School leaders support	• School leaders need to provide time for teachers to plan, analyse and reflect together on the process and outcome of trying new things. • Encourage risk taking.
Collaboration	• Learning how to learn from close observations of learning and teaching exchanges.
Structured dialogue	• Structured dialogue rooted in evidence from trying things out with pupils that disturb the status quo.
Ambitious goals	• Ambitious goals may be mandated externally provided there is a strong element of peer support through which instructions from others can be interpreted from professional learners' own pupils.

Source: Author's own work based on Bell et al. (2010), Cordingley et al. (2007), Cordingley et al. (2005a), Cordingley et al. (2005b), Cordingley et al. (2003), Timperley et al. (2007).

As one systematic review of teachers' use of research demonstrated, the quality of coaching, mentoring or support is a significant accelerator or inhibitor (Bell et al., 2010). Some HE colleagues involved in supporting teacher research do so from a love of teacher enquiry but lack specialist knowledge in the content of the area of research being explored by the teachers they are supporting. Some mentors in particular fields have specialist knowledge and expertise but lack knowledge and experience in applying research techniques to the demands of enquiry within busy, dynamic, messy school environments. Other HE colleagues with research expertise that could accelerate teacher engagement with knowledge and evidence are used to working to more extended timescales and are unable to provide the brisk project management and business-like support that teachers and schools need to mesh enquiry activities with the rhythm of day to day school life.

Effective support for teacher engagement with evidence calls for an unusual combination of skills that is usually more easily found through accessing a network of colleagues rather than through bilateral relationships. Such networks require structural encouragement and support from, for example local districts and universities if they are to flourish and grow quickly. In England, funding for school based research consortia and Networked Learning Communities were two successive, early national initiatives that had some success in building a networked infrastructure for the support of teacher use of research such as the Networked Learning Communities programme (Earl and Katz, 2005). Teaching Schools are a concept in more recent initiatives seeking to achieve similar momentum within a more self-directing system (Sebba, Kent and Tregenza, 2012; Hargreaves, 2012).

Box 7.2. Supporting teachers to engage with research

What then does this abstract collection of key characteristics look like on the ground? One interesting example of a professionally-driven approach to capacity building through engaging teachers in and with research was launched in the early part of the twenty first century by the National Union of Teachers (NUT), the biggest English Professional Association at the time. The NUT had in fact been the original sponsor of the first of the systematic reviews of evidence about what makes a difference for teachers and for pupils, and sought to establish and model professional development and capacity building in a way that aligned closely with best evidence. Their "teacher2teacher" CPD programme involved pairs of teachers in working together on a sustained period to develop and evaluate emerging practice-based on intense working with leading edge researchers over twenty-four hours. The topics for "teacher2teacher" programmes arose from requests for NUT members, the views of NUT policy officers about system level issues causing teachers concern and the views if their substantial body of members who were also school leaders. Leading edge researchers were identified and recruited on the basis of their research publications and after considerable desk research and consultation across NUT's extensive network of researchers who from whom they had commissioned research. These included, for example, members of the original Black and Wiliam research, David Wray, one of the authors of The Effective Teachers of Literacy report for the Teacher training Agency (Wray et al., 2000) and Robert Fisher, author of a number of studies on the use of thinking skills in primary schools (Fisher, 2013).

Box 7.2. Supporting teachers to engage with research *(cont.)*

During the initial twenty-four-hour residential workshops teachers were immersed in illustrations of new approaches, in experimenting with tools and resources that nest them in classroom practices and in planning to experiment with them, over three cycles of experimentation and reflection that spanned roughly twelve weeks. During the initial residential, the teachers learned about the evidence about collaborative coaching and built structured, formal Learning Agreements. The objective was to shape their expectations of how they would work, the evidence they would collect about how their learning connected with pupil learning and the ways teachers would support each other's, sometimes quite different, projects. After approximately twelve weeks the teachers came together for another intense workshop focused on analysing how each other's experiments had worked, exploring together changes in pupil learning and work, photographs and videos of lessons, lesson plans and changes in their thinking and understanding. This reflection and analysis was facilitated by the original specialists.

The final stage of the programme involved the teachers planning how to translate their own learning into learning experiences for their colleagues, role-playing the initial stages and considering how they would be able to a) continue their own learning as part of the process of supporting others and b) how they would know their own and their colleagues' learning had been successful. Some of these teachers went on to write up their learning experiences and others used this embedded form of engagement with and in research as a springboard for embarking on more explicit research for doctorate and masters programmes. NUT itself then established a series of scholarship projects focused on key NUT priorities such as Thinking Skills and improving the quality of talk which enabled teachers to progress to a more formal mode of engagement with and in research and several other "graduates" of these programmes subsequently supported and promoted teacher engagement in and with research by, for example, and serving as members of teacher research groups including the National Teacher Research Panel.

During the first ten years, NUT ran these programmes for between eight and twelve different groups of teachers and focused on a wide range of different priorities. It is still continuing over a decade since it started and in times of austerity; in this instance in relation to development education.

Implications for governance in complex systems

This second part of the chapter takes a closer look at how teachers' engaging with and in research is related to governance in more general terms. Regarding effective capacity building through improving knowledge use and transfer, the evidence lets us identify three broad fields in relation to governance where support processes and structures are likely to be needed:

Encouraging teachers and managing risk: generating confidence in and a thirst for high impact approaches emerging from research. To build capacity, teachers need to be encouraged not to shy away from the difficult but important issues rather than reaching for undemanding issues close at hand. This means encouraging teachers to take risks and calls for governance processes able to manage the risks of failure (see also Burns and Blanchenay, Chapter 10).

Facilitating access: the supply of accessible and usable summaries of research that offer multiple entry points and support for teacher engagement in and with research, a pathway through the illustrations of evidence about high impact approaches at work in classrooms, the tools and protocols for using them and the clear explanations of the underpinning theory or rationale on which such depth depends.

Assuring quality: Quality assurance pertains to processes and systems for promoting depth and assuring quality in support for teacher engagement in and with research.

Encouraging teachers

In England, teachers interested in research and evidence grew to the point that in 2010, almost 40% of teachers reported formal engagement in their own research or with the research of others during the previous 12 months. This is a remarkable change that arose from a mix of top down policy leadership, sideways-on support from the General Teaching Council, the professional associations and local authorities and bottom-up demand from teachers and school leaders (Cordingley, 2010).

Top-down policy making in England, for example through national guidance about the curriculum and national teaching and learning strategies, became increasingly explicit about the way evidence was informing policy over the course of the 2000-10 decade. This was in turn reflected in the systematic embedding of evidence in the support materials and CPD different policy agencies such as the Qualification and Curriculum Development Agency, the Teacher Development Agency and the Department for Education plus the General Teaching Council offered to schools. It is also reflected in the decisions of the new Coalition government about specific interventions such as synthetic phonics, even though in the main their policies have advocated significant reductions in the level of explicit prescription to, or central support for, schools in an effort to increase their autonomy, self-direction and accountability.

But much of the effort to increase demand was aimed to generate interest "from the bottom up". Grants for flagship teacher research champions, encouragement of school-to-school networking via engagement with research, and embedding understanding of the role of research in effective CPD all helped to increase demand. So too did giving a high profile to examples of effective engagement with research. For example, in England the National Teacher Research Panel has played an important role in encouraging teachers to engage in their own research and with the research of others. The conferences comprised workshops run by teachers whose research had been peer-reviewed by the Panel against criteria relating both to the quality of the research and to its relevance and usability. Successful applicants were coached on how to summarise their research in ways that would contribute to other teachers' learning, and on how to design interactive workshops. At each conference, some 40-60 teacher researchers showcase excellence in engaging in their own research or with the research of others; the results are made available by the Panel's popular website[2] and are also frequently used to illustrate larger scale academic research via, for example, the Research for Teachers resources[3]. Panel members also used the summaries to support local and regional research networks and to run local and regional teacher research conferences.

Teachers' professional associations also played an important role in championing and highlighting teacher engagement in and with research. In England, the National Union of Teachers played an important role by funding research reviews, role modelling research-based CPD and funding teacher scholarships (Box 7.2). The Association of Teachers and Lecturers (ATL) similarly sponsors and encourages teacher engagement in study at master's level that includes carrying out research. More recently the formally designated Teaching Schools have been given responsibility for leading 6 strands of development activity across an Alliance of between 5 and 40 schools, and research and development is one (albeit the least well established) of the key strands.

Managing risk

A cornerstone for securing an effective flow of empirically supported knowledge is ensuring that what is offered speaks to "wicked issues". That is, to teachers' aspirations and concerns for their pupils' learning. Huberman (1993) argued that the research community should be collecting and analysing the questions teachers pose of their practice during CPD and enquiry planning to shape the education research agenda. Using challenging and complex approaches such as Assessment for Learning in busy classrooms means unlearning established safety routines and control mechanisms. This is sustained, hard, emotional as well as intellectual work that needs to overcome considerable practical and operational obstacles. Obstacles include, for example, the pressure to do things quickly; new approaches take extra planning and use more classroom time in the early stages. They also affect, and possibly undermine, existing tried and tested, routinised practices in unpredictable ways.

There is also a risk that misunderstanding the underpinning rationale for new approaches could lead, in the process of adapting them for particular students, to unintended consequences, and inadvertently removing its core features. These challenges lead to a number of important practical considerations for those seeking to grow capacity in this way. Developing new or enhanced research informed approaches also means providing and then steadily removing scaffolding, for example through tools that limit some of the demands on teachers' attention, or help them explore the connections between new strategies and pupils' learning in progressive waves whilst also ensuring that current orthodoxies and assumptions are challenged.

This means ensuring that evidence about how pupils are responding to new approaches needs to be built very explicitly into the development process to ensure that risks to students are identified and managed. It also reinforces the importance of selecting approaches that have more than local, anecdotal evidence to suggest that the bumpiness of early experiments with new approaches will lead to benefits that outweigh the risks.

Facilitating access by providing the right tools

Tools are important in supporting use of evidence at scale because tools enable leaders to secure consistency and coherence in the way leadership policies are applied (Robinson, Hohepa and Lloyd, 2009). Tools were an explicit feature of policies for promoting evidence-informed practice in England from 2000-10, through a number of parallel attempts to broker and mediate access to research in user-friendly forms (Table 7.2).

Table 7.2. Examples of tools to facilitate access to research

Example	Description
Short presentation-type summaries	• Short presentation-type summaries of high quality research that would take a teacher just 2.5 minutes to read but which also provide reflective questions that would encourage teachers to explore further.
"Research tasters"	• Micro enquiry tools called "research tasters" comprising a distillation of key and intriguing research findings in about 50 words. • A mini evidence collection and recoding activity teachers can use to collect evidence from pupils. • Reflective questions for exploring the evidence from their pupils. • Recommendations about experimenting with the approach outlined in the nugget and using the enquiry tool to continue to collect evidence about how pupils are responding. • Links to further information.
Research papers digests	• Medium detailed digests of high quality research papers that are relevant to practice using a standard format. • Derived from research observation of teachers exploring a range of research papers and discussion with them about features they find most helpful.
Larger summaries	• 5 000 word summaries of the highest quality, large scale studies and a small number of well-tested research, linking every academic finding to a good quality teacher-researched case study.
Anthologies	• Anthologies for specific sub groups of teachers, e.g. newly qualified teachers or teachers with a particular interest in equalities, in which themes emerging across the studies were illustrated by "research tasters".
Research magazines	• Research magazines for teacher users of research, each tailored to the constituencies and policy briefs of different government agencies.

Source: Author's compilation based on policies in place in England (United Kingdom) 2000-10.

Underneath this array of resources designed to improve the supply of research evidence to teachers sit different levels of engagement with the research of others. There is an entry-level need to raise awareness of the range of potential benefits that the research evidence base has to offer, and an associated need to secure understanding of the core facts and issues revealed by particular pieces of research to enable teachers and schools to consider their relevance. At a more intermediate level there is a need to encourage teachers to experiment with approaches highlighted by research in the context of evidence about their own pupils and context and to interpret and refine approaches for that context. At the most sophisticated level there is a need to provide access not just to the evidence about an intervention but also to the underpinning principles so that teachers can develop a practical theory or rationale for their work and to inform the adaptations they make as they embed new approaches in range of different contexts. Tools are also crucial for effectiveness because they help to make teacher learning more visible and so enable better understanding of demands that new approaches are making on teachers (Robinson, Hohepa and Lloyd, 2009). The tacit nature of teachers' professional knowledge is, as this chapter describes, a major influence on their use of other forms of knowledge.

It is worth considering how different groups of teachers conceptualise their own knowledge in this context. Effective teachers internalise complex knowledge and skills to the point where they are able to use their conscious attention to focus on the particular learners they are working with; and thus to the point where they are barely aware such skills are put to use. Such teachers often describe much of their skilled, dynamic and

complex practice as simply a matter of "common sense". Unfortunately herein lies an important source of potential confusion. Because new and less skilled teachers also describe much of their practice as common sense, they cling onto published materials that are familiar and feel very manageable but which may be very weak, or onto tried and tested regimes whose main function is to control behaviour rather than to enable learning. Such teachers also see their practice as "common sense"; and may well believe they are making common cause with more developed colleagues in doing so, even though they may end up using this notion of common sense knowledge to justify resistance to improvement projects or to the emotional and intellectual costs necessary to achieve significant improvements in teaching and learning. The vernacular "common sense" with such diverse roots thus obscures almost diametrically opposed stances and may fuel resistance to change except where schools have established an effective professional learning environment whose role is to engage all teachers in challenging practice and orthodoxies together using, for example, action research, lesson study of evidence based collaborative coaching as tools for developing new professional knowledge.

Assuring quality

The challenges in England have not simply been practical ones. Early support for building teacher capacity through national research grants triggered an outbreak of methodological wars, first in the Times Education Supplement and later in research journals (Hammersley, 1997; Gorard, 2001). Contestation focused on about whether teacher research is real research and about whether or not teacher research should only focus upon evidence from the profession's own practise.

Interestingly, some ten years later, the Practitioner Use of Research Review described above brought evidence to attention that teachers were engaging with evidence from practise in their own classrooms and those of their colleagues and with evidence from larger scale, academic studies (Bell et al., 2010). But the debate about the validity of teacher research for informing others' practice still rumbles on, as the review of education research by Ben Goldacre (2013) for the new UK coalition government showed, by advocating strongly that teachers should not be undertaking their own research but looking for and participating in researcher led randomised control trials.

In England the general belief is that the quickest way for activities to be embedded at scale across the system is for them to be included in the OfSTED[4] inspection framework. The most recent revisions to that framework do in fact place considerable emphasis on continuing professional learning that is properly connected to pupil learning. It remains to be seen how many schools and inspectors make the link between that and engagement in and or with knowledge and evidence research but if they do that is likely to significantly increase demands for research tools, resources and activities as a core strand of school improvement. Recent changes to the OfSTED inspection framework to increase the validity of judgements that were previously made about teacher quality on the basis of 20 minute classroom observation shows that OfSTED too are having to pay increasing attention to the disciplines of research. Challenges from academic commentators on the reliability of OfSTED judgments (Stewart, 2013) and reflective responses from Mike Cladingbowl (2014), then Director of policy at OFSTED, have brought requirements that inspectors should triangulate evidence from observations from 20 minute visits to lessons, with evidence from pupils' work books and discussions with them.

Recently the UK Government has launched a "Close the Gap, Test and Learn programme" as a centrally designed but locally led Research and Development (R&D)

initiative focused on closing gaps for vulnerable pupils. This positions R&D leads in Teaching School Alliances as, in effect, local managers for trialling, on a randomised basis, seven interventions across over 750 schools. Not only are R&D leads responsible for recruiting schools, and explaining the nature and purpose of randomisation, they also have a role in helping them with testing and encouraging qualitative research about, for example, fidelity, alongside the quantitative on-line assessments. At the time of writing, the results of this radical and large scale programme to promote research and evidence informed practice are still pending. However, it is already clear that putting teachers in a leadership role around R&D has helped greatly both with recruitment of schools to a trial and with the retention of control schools. This larger scale approach to engendering and supporting teacher engagement with evidence as a means of aligning knowledge from both practice and research and building local capacity for improvement has certainly created energy and momentum.

Conclusion

This chapter suggests that using knowledge and research and capacity building for evidence informed professionalism within complex governance systems are learning problems. What we know already about supporting the learning of young people has much to tell us about how we support the learning of teachers (Cordingley, 2008). If, as argued here and in the reviews of evidence about effective CPD outlined above, school leaders need to approach supporting their staff as though the staff were their class, perhaps policy makers would find it helpful to consider structures and policies as though they were the improvement curriculum for the education system and to approach the ways these are enacted as system level pedagogy?

In the task of developing systems to underpin such work, it might also be useful to conclude by listing some of the challenges encountered as the research and evidence-informed policies in England unfolded, and strategies adopted for tackling them. This might provide a reasonable springboard for considering how governance can be used to develop such capacity and benefit from it. The debates in the late 20th century and early 21st about the role of teacher engagement in and with research compared to the role and quality of large-scale research were heated. Noticing the distinction between teacher engagement with the research of others and in their own research and the importance of both was helpful in positioning teachers as having an interest in connecting the two. Another pathway through the opposing views was created by distinguishing between:

- The importance of the generation and recognition of large scale research and evidence as important for deciding whether to pursue an approach as a policy that is to be imposed on others.
- The importance of the collection, analysis and interpretation of fine grained, relevant, triangulated qualitative evidence at scale. Focussing on the processes underpinning findings about the impact of different approaches can help to shape efforts to test and replicate high leverage approaches.
- The potential of smaller scale and / or qualitative evidence generated by practitioners as they test out and contextualise larger scale findings and responses to local challenges.

This last element is key to helping teachers feel that such efforts are possible in their own context. Teachers' (quality assured) systematic accounts of development experiences seem to be especially compelling to their colleagues, perhaps because they help them develop a sense of collective efficacy. In this context illustrative research by teachers

geared to improving their own practice and inspiring and informing similar improvements for others has an important part to play in connecting generalisable knowledge and evidence with practice based knowledge and practitioners' aspirations for their pupils.

Making and exploring these distinctions has helped all the strategic players (policy makers, researchers, teacher organisations, teacher educators, policy makers and school leaders) see that each had an important contribution to make to pre-empting over-polarisation of the lines of argument or a hardening of the different interests and perspectives involved in connecting evidence and practice.

Further insights that emerged in the English context included:

- *Teachers' individual and collective contributions to research informed practise need to be appreciated.* The establishment of the English National Teacher Research Panel (ENTRP), comprising a group of 15 teachers able to provide extensive evidence about their engagement in their own research and their use of others' research was helpful here.[5]

- *Teachers should be helped to develop the confidence and skills to analyse and evaluate the relevance of research evidence whatever its provenance.* One early strategy that bore some fruit was involving expert teacher researchers from the ENTRP in developing a framework for exploring the quality of a wide range of knowledge and using this to peer review and model excellence in teacher research evidence to increase "research literacy" across the profession (ENTRP and Cordingley 2003). These guidelines were used explicitly by the panel to attract and select high quality teacher research for their biennial conferences, to identify larger scale studies to inform the Panel's work and to inform teachers' contributions to the many research advisory groups on which they sat. Funding of such panels and for teachers to participate in research advisory groups, to peer review teachers' own research and to convene conferences of quality assured teacher research at local, regional or national levels could play an important role here. General Teaching Councils are increasingly getting involved in such work and it also seems likely, at the time of writing, to feature in the role and development work of an *English National College of Teachers.*

Developing teacher access to high quality, systematic and technical reviews of research in areas where teachers have concerns for pupils and where studies have been extensive. Teachers, like policy makers, have little time to trace through the sometimes byzantine often erratic pathways between partial or small scale, sometimes conflicting studies. Nor do most of them have access to the expensive library archives available to Universities. In addition to better access to teacher and policy friendly summaries of individual study findings, teachers need access to systematic research synthesises.

At the end of the 20th century Black and Wiliam's seminal work on Assessment for Learning illustrated the art form and the English government set up a centre to develop and quality assure such reviews, the Evidence for Policy and Practice Centre (EPPI)[6], to build on this (Black and Wiliam, 1998). Subsequent reviewing methodologies pushed the boundaries of such reviews further to the point where the excellent and rigorous New Zealand Best Evidence Syntheses gave teachers and parents a direct role in signing off review protocols and findings as having the potential to improve the quality of teaching and learning. More recently the Hattie review and synthesis of the effects of different interventions has become renown amongst both policy makers and practitioners (Hattie, 2009).

National Knowledge centres are also emerging, for example, in Belgium, Norway, and Denmark as a means of developing more coherent, national approaches to use of evidence and knowledge. Knowledge services are growing rapidly. An example from England is the "Sutton Trust Toolkit" which is promoted and funded by the Education Endowment Foundation[7]. This web based system of evidence assessment for particular interventions is based on a randomised trial approach to knowledge mobilisation at scale. It will be important to the contribution of knowledge services to governance, capacity building and enhancing students' life chances to ensure that teachers will take ownership and have a stake in the resulting structures. The English Government's decision to link this tool-kit to the evaluation of how schools are deploying government funds for vulnerable students at the same time as promoting Research and Development via Teaching School Alliances is an interesting early experiment.

Notes

1. Post Graduate Professional Development Programme funded by the Teacher Training Agency (CUREE, 2009), the Best Practice Research Scholarships (Street and Temperley, 2005) funded directly by the Department for Education and skills (DfES), the Networked Learning Communities programme (Earl and Katz, 2005) and the research associate programmes funded by the National College for School Leadership.
2. Available at www.ntrp.org.uk/.
3. Available at www.tla.ac.uk/site/Pages/RfT.aspx.
4. Office for Standards in Education, Children's Services, and Skills, www.gov.uk/government/organisations/ofsted.
5. Evidence of their lasting legacy can be found at www.ntrp.org.uk.
6. Evidence for Policy and Practice Centre (EPPI), www.eppi.ioe.ac.uk.
7. Education Endowment Foundation, www.educationendowmentfoundation.org.uk/.

References

Bell, M. et al. (2010), *Report of Professional Practitioner Use of Research Review: Practitioner Engagement in and/or with Research*, CUREE, GTCE, LSIS & NTRP, Coventry, www.curee-paccts.com/node/2303.

Black, P. and D. Wiliam. (1998), *Inside the Black Box: Raising Standards through Classroom Assessment*, School of Education, King's College, London.

Cladingbowl, M. (2014), *Why I Want to Try Inspecting Without Grading Teaching in Each Individual Lesson*, Ofsted, London, http://webarchive.nationalarchives.gov.uk/20141124154759/http://www.ofsted.gov.uk/resources/why-i-want-try-inspecting-without-grading-teaching-each-individual-lesson.

Cordingley, P. (2010), "Stepping stones, bridges and scaffolding: Effective tools, artefacts and professional learning processes for research use", paper presented at American Educational Research Association (AERA) Conference, Denver, 30 April-4 May 2010.

Cordingley, P. (2008), *Sauce for the Goose: Learning Entitlements that Work for Teachers as Well as for Their Pupils*, CUREE, Coventry.

Cordingley P. et al. (2005a), "The impact of collaborative CPD on classroom teaching and learning. Review: What do teacher impact data tell us about collaborative CPD?", in *Research Evidence in Education Library*, EPPI-Centre, Social Science Research Unit, Institute of Education, University of London, http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=395.

Cordingley P. et al. (2005b), "The impact of collaborative continuing professional development (CPD) on classroom teaching and learning. Review: How do collaborative and sustained CPD and sustained but not collaborative CPD affect teaching and learning?", in *Research Evidence in Education Library*, EPPI-Centre, Social Science Research Unit, Institute of Education, University of London, www.eppi.ioe.ac.uk/cms/Default.aspx?tabid=392&language=en-US.

Cordingley P. et al. (2007), "What do specialists do in CPD programmes for which there is evidence of positive outcomes for pupils and teachers?", in *Research Evidence in Education Library*, EPPI-Centre, Social Science Research Unit, Institute of Education, University of London, www.eppi.ioe.ac.uk/cms/Default.aspx?tabid=2275.

Cordingley P. et al. (2003), "The impact of collaborative CPD on classroom teaching and learning", in *Research Evidence in Education Library*, EPPI-Centre, Social Science Research Unit, Institute of Education, University of London, www.eppi.ioe.ac.uk/cms/Default.aspx?tabid=133&language=en-US.

Cordingley, P. and B. Crisp (2014), "The challenges and opportunities involved in designing large-scale national "RCT-like" programmes in education", paper presented at British Educational Research Association Conference, London, 23-25 September 2014.

CUREE (2009), *Postgraduate Professional Development (PPD) Programme: Quality Assurance Strand: Research Report Year 3*, CUREE, Coventry.

Earl, L. and S. Katz (2005), *What Makes a Network a Learning Network?*, National College of School Leadership, Cranfield.

Fisher, R. (2013), *Teaching Thinking: Philosophical Enquiry in the Classroom*, Bloomsbury Academic, London.

Goldacre, B. (2013), *Building Evidence into Education*, Department for Education, London, www.media.education.gov.uk/assets/files/pdf/b/ben%20goldacre%20paper.pdf.

Gorard, S. (2001), "A changing climate for educational research? The role of research capacity-building", paper presented at British Educational Research Association, Leeds, 13-15 September 2001.

Hammersley**,** M. (1997), "Educational research and teaching: A response to David Hargreaves' TTA lecture", *British Educational Research Journal,* Vol. 23/2, pp.141-161.

Hargreaves, D.H. (2012), *A Self-Improving School System: Towards Maturity*, National College for School Leadership, Nottingham, www.dera.ioe.ac.uk/15804/1/a-self-improving-school-system-towards-maturity.pdf.

Hattie, J.C. (2009), *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*, Routledge, London.

Huberman, M. (1993), "Changing minds: the dissemination of research and its effects on practice and theory", in C. Day, J. Calderhead and P. Denicolo (eds.), *Research in Teacher Thinking: understanding professional development*, Falmer, London.

Marshall B. and M.J. Drummond (2006), "How teachers engage with Assessment for Learning: lessons from the classroom", *Research Papers in Education,* Vol. 21/2, pp. 133-149.

Mourshed, M., C. Chijiokeand and M. Barber (2010), *How the World's Most Improved School Systems Keep Getting Better*, McKinsey & Company, New York.

NTRP (National Teacher Research Panel) and P. Cordingley (2003), "Encouraging and supporting CPD in making use of research: Guidelines from the National Teacher Research Panel", Paper presented at British Educational Research Association Conference, Edinburgh, 11-13 September 2003.

OECD (2015), *Schooling Redesigned: Towards Innovative Learning Systems*, Educational Research and Innovation, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264245914-en.

Robinson, V., M. Hohepa and C. Lloyd (2009), *School Leadership and Student Outcomes: Identifying What Works and Why*: *Best Evidence Synthesis Iteration (BES)*, New Zealand Ministry of Education, Wellington.

Sebba, J., P. Kent and J. Tregenza (2012), *Joint Practice Development: What Does the Evidence Suggest are Effective Approaches?*, National College of School Leadership, Nottingham.

Stern, C. (1949), *Children Discover Arithmetic*, Harper and Rowe, New York.

Stewart, W. (2013), "Ofsted's approach 'is not backed by research'", *TES Magazine*, London, www.tes.co.uk/article.aspx?storycode=6356566.

Street, H. and J. Temperley (2005), *Improving Schools Through Collaborative Enquiry*, Continuum International Publishing Group, London.

Tacon, R. and T. Wing (2004), *A Multi-Sensory Approach to Teaching Mental Arithmetic*, National Teacher Research Panel, Coventry, www.ntrp.org.uk/node/62.

Timperley, H. et al. (2007), *Teacher Professional Learning and Development: Best Evidence Synthesis Iteration*, New Zealand Ministry of Education, Wellington.

Wray, D. et al. (2000), "The teaching practices of effective teachers of literacy", *Educational Review*, Vol. 52/1, pp. 75-84.

PART 4.

COMPLEXITY IN POLICY MAKING: THINKING STRATEGICALLY

Chapter 8.

Policy experimentation in complex education systems

Patrick Blanchenay, *University of Toronto*

Tracey Burns, *Centre for Educational Research and Innovation, OECD*

Complexity is increasing in education – in governance arrangements, in the numbers of stakeholders and in the availability and use of evaluation and other accountability data. These changes call for moving away from a traditional policy cycle towards one which can evolve and adapt with our systems in order to govern them effectively. One tool of this new kind of governance is policy experimentation.

This chapter suggests avenues to make experimentation a more effective instrument for policy making in a complex environment, and demonstrates that a tension exists between properly evaluating the effects of narrowly-focused experiments and translating these results into the broader network in which every stakeholder is embedded. It suggests that a good balance can be struck by experimenting at a suitable scale, and moving towards what is called ecosystem experimentation.

Introduction

The increased devolution of responsibilities and tasks from central governments to lower levels of government and local authorities, combined with a trend towards increasing accountability, has contributed to make education systems in the OECD countries more intricate and more complex. This trend has happened both vertically, with the interaction of different levels of governance, as well as horizontally, with increased involvement of a wider range of actors in all processes.

This increased complexity – in governance arrangements, in the numbers of stakeholders, and in the availability and use of evaluation and accountability data – calls for a new approach to governance. Education systems are in fact *complex systems* – that is, networks of interdependently linked actors whose actions affect all other actors, and which evolve, adapt, and reorganise themselves. Complex systems do not work in a linear manner but rather exhibit a series of well-defined characteristics: tipping points, feedback loops, path dependence and sensibility to local contexts (Bryne, 1998).[1]

Understanding complexity is an important point for policy making and governance in general, as complex systems cannot be successfully governed or steered with simple, linear mechanisms. In complex environments, strategies must be developed that take into account the dynamics and interdependency of the system. Simply devolving power to local authorities will not improve the functioning of the system unless it is also accompanied by attention to the connections and interactivity present, and space is made to facilitate and use the constant feedback that is required to guide complex systems. However, in complex environments such as education in which a multitude of actors are collaborating through formal and informal channels, the sheer amount of feedback and interactivity can seem impossible to navigate effectively.

Policy experimentation has been suggested as one strategy for dealing with such complexity. Policy experimentation can be defined as "a purposeful and coordinated activity geared to producing novel policy options that are injected into official policymaking and then replicated on a larger scale" (Heilmann, 2008b). In practice this implies the deliberate implementation of a new programme or practice on a small scale, targeting a selected number of schools or districts, with the intention of evaluating the effectiveness and possible scaling up to a wider level if effectiveness is demonstrated.

The present chapter provides a rationale for the use of policy experimentation in the context of complex education systems, and shows under what circumstances it could be useful. First, the chapter outlines broad governance challenges posed by complexity. It then defines more precisely what is meant by experimentation, what we can learn from it in a complex environment, and places experimentation as a useful tool in the debate between big package interventions and focused incremental reforms. In a third section, the chapter shows how experimentation can be tailored to account for complexity by choosing a suitable scale that we call "ecosystem experimentation", and by adopting the policy cycle to account for such complexity. It then looks at the risks involved with experimentation, and highlights the necessity of building educational systems capable of taking risks and can learn from both success and (importantly) failure.

Finally, the chapter suggests avenues to make experimentation a more effective tool for policy making in a complex environment. A tension exists between properly evaluating the effects of narrowly focused experiments and translating these results into the broader network in which every stakeholder is embedded. It suggests that a good

balance can be achieved by experimenting at a suitable scale, and moving towards what is called *ecosystem experimentation*.

Governance challenges in a complex environment

Across the OECD, education governance has in general moved away from hierarchical governance systems towards more complex environments in which a multitude of actors collaborate through formal and informal channels. In such contexts, the successes and failures of students and schools depend on a multitude of interdependent actors, who all play a part in moving the system forward. This interdependence poses several challenges for the governance of education systems that are critical for this discussion:

1. The traditional linear approach to policy-making may not be best suited to operate in such an environment.
2. Successful governance requires the co-operation of actors with different motivations, outlooks and time horizons.
3. Information is more plentiful and also more scattered, both in its production and its consumption.

The traditional policy cycle is inadequate

The complexity of educational systems means that the traditional policy cycle is not able to capture the interplay and dynamics characteristic of modern arrangements. This is not radical or startling news: in fact, there have been numerous critics describing the inadequacy of the traditional policy cycle, and not just in education. Clay and Schaffer (1984) made this argument in relation to agricultural policy 30 years ago, and these comments have been broadened and deepened in almost all public sectors since. Hallsworth et al. (2011: 38-44) provide a detailed breakdown of why this is so:

- "Policy making does not occur in distinct stages". Instead, problems and potential solutions often emerge together, rather than sequentially.
- "Policies need to be designed" properly to reach their goals, but it is hard to design something perfectly without a trial or implementation attempt to refine it. This problem is magnified by the fact that feedback loops in a complex system may amplify the unpredictability present and the whole set-up of links in the system shifts as the experiment progresses, thereby creating potentially unintended consequences.
- "Policy making is often determined by events". Politics and media attention are important forces affecting potentially powerful stakeholders; this may sometimes lead to decisions taken for political reasons, often with an emphasis on short-term results. In the age of social media, real time examples of the power of different stakeholders (for example parents or students) in changing the terms of a debate or calling into question the adequacy of a policy response to a particular problem can be powerful forces in politics and policy.
- "The effects of policies are often indirect, diffuse and may take time to appear". As already mentioned, inference is made more difficult in a complex environment, which in turn implies that the evaluation of certain policies or experiments is challenging and might not readily translate into the creation of new

initiatives (or at least, not in the timeline required by the traditional policy process or government mandates, that is, two to four years).

While Hallsworth and his colleagues wrote this chapter with the UK policy cycle in mind, their insights apply to many other national contexts and are clearly pertinent for education systems.

The inadequacy of the standard policy cycle in a complex educational environment is further magnified by the fact that seemingly similar contexts may have very different dynamics and therefore be affected differently by the same policy. For instance, disclosing information about school performance might have very different impacts on a school that is thriving, as opposed to a school that struggles to attract well performing students. Whereas the traditional policy cycle usually operates on the back of a top-down framework of policy making, complexity instead requires policy cycles to operate closely enough to the local level so that policies can be tailored to idiosyncratic contexts, rather than follow a one-size-fits-all approach.

Two other elements contribute to making governance in complex environments challenging: a) the differences in objectives and time horizons of the various stakeholders and b) the informational challenges that arise as a function of the increased availability of data and information more generally. These will be discussed briefly in turn in the next sections.

Differences in objectives and time horizons

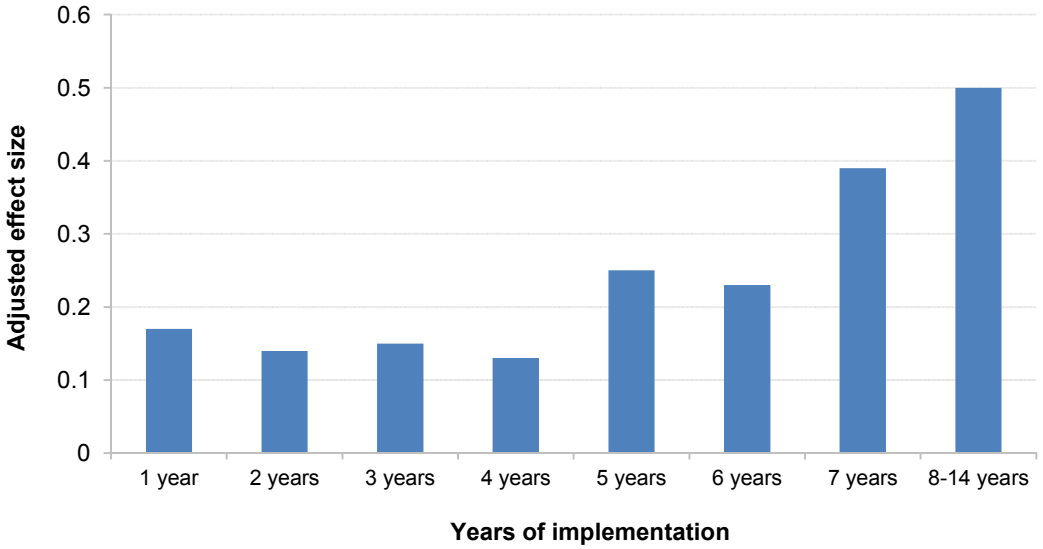
Two combined trends in the governance of education systems, and indeed many of the public sectors (e.g. health, justice, etc.) have jointly contributed to put stakeholders with very different perspectives at the heart of the decision making processes; on the one hand that of devolution of power towards lower more local levels of decision making, and on the other hand that of increased accountability and widened participation in the decision process. These trends have implied that education systems now tend to encompass more stakeholders than before: decision nodes have been more scattered both vertically within education professionals, and horizontally towards surrounding communities and other actors. The functioning of education systems has more than ever become the product of joint actions by students, educators, parents, school directors, civil servants and elected officials at various levels of government.

With this diversity of actors also comes diversity in expectations. For example, elected officials have to operate on shorter time scales than civil servants, teachers, parents and students. This can give rise to different policy preferences when choosing interventions, particularly as they relate to time and level of risk required. For example, quick-effect changes (e.g. providing students with electronic tablets) might become more appealing to elected officials as elections loom closer, while parents may favour longer-term less risky changes (e.g. reinforcing the teaching staff) and researchers may prefer more risky longer-term experiments (e.g. teaching a new reading method). Paradoxically, moving from appointed to elected officials as a way to increase local accountability in the education system, for instance in school boards, might result in an undesirable preference for more visible short-term solutions from those officials, given the requirements of the electoral cycle. The introduction of electoral accountability can also induce a succession of short-term reforms that may induce "reform fatigue" among the stakeholders impacted.

This preference for policies that yield effects in a shorter time span can sit uncomfortably in the realm of education, where policies may take a long time to take

effects. Indeed, one longitudinal model of comprehensive school reform suggests that the strongest effects are seen 8 to 14 years after a reform is begun, as shown in Figure 8.1 (Borman et al., 2003). This is due to a number of factors. In the realm of school choice, for example, Waslander et al. (2010) point out that reforms can take a long time to bear effects, both in terms of when the policy comes into practice and when parents act upon it. They also emphasise that time is not unidimensional but in fact interacts in highly complex ways with the actors and the context. In fact, short-term effects can be different from long-term effects because, using the school choice example again, certain types of parents might be quicker to react to the policy change than others. This implies that the kinds of impacts that are observed in the short term may be qualitatively different, both in type and intensity, than those that develop over a longer timespan. This is a particularly relevant observation when combined with the knowledge that policies do generally require quite some time to take full effect. Such conclusions show that it is sometimes necessary to leave to reforms a longer time span than the natural time span induced by electoral cycles.

Figure 8.1. Effect sizes of compulsory school reform in the years following implementation



Note: Effect sizes based on meta-analysis of studies pertaining to the impact of comprehensive school reform on student achievement in the United States.

Source: Borman et al. (2003).

This difference in time scales is compounded by a difference in objectives: officials, in particular elected officials, will favour finding "what works" while researchers may focus on understanding "why it works". Although ideally these two approaches should combine, in reality this is not always possible, in particular when facing budgetary constraints and thus the necessity of prioritising various issues.

Fractionalisation of decision making generates informational challenges

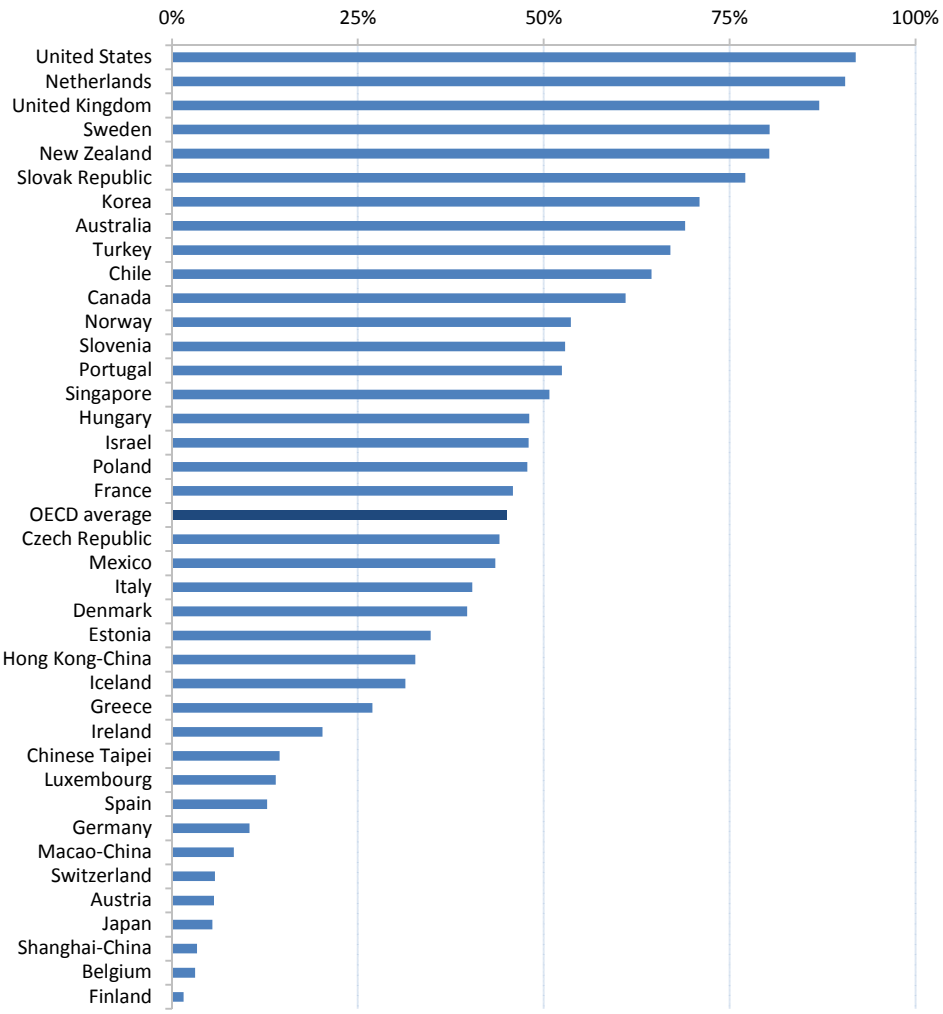
As already argued, education systems are becoming more complex through increasing devolution of decision making and the involvement of a wider range of stakeholders. They are also becoming more complex due to a greater availability of data on educational performance and other system factors that are relevant for the decision-making process.

The increase in the availability of information is one of the most dramatic transformations in our education systems in the last two decades, fuelled by two concurrent trends. The first is the rise of standardised tests (both national and international, for example PISA) and the resulting explosion of available evidence and greater emphasis on testing and assessment. The second is the increased access to information via the Internet and other technologies, which has enabled a multitude of actors to bring their own informed opinions to the discussion.

According to PISA 2012, on average across the OECD 45% of students are in schools whose principals report that achievement data are posted publicly. But this average belies a very wide distribution: in the United States, the Netherlands, the United Kingdom, Sweden and New Zealand, over 80% of students attend such schools. At the other end of the scale, less than four per cent of students attend schools where achievement data are posted publicly in Belgium, Finland and Shanghai (China) (OECD, 2014, see Figure 8.2).

Figure 8.2. Use of school achievement data for accountability purposes

Percentage of students in schools where achievement data are posted publicly



Note: Included in graph: all OECD countries and the top 5 PISA performers in mathematics.

Source: OECD, PISA 2012 Database, Table IV.4.13. http://dx.doi.org/10.1787/888932957346.

This abundance of data can pose challenges in an environment where different actors, with different needs, cultures and perspectives need to share information with each other. First of all, the availability of large amounts of data must not be confounded with having a full understanding of any given situation. Current data collections omit important (and potentially explanatory) variables on issues as diverse as student well-being, the role of non-cognitive skills in student achievement and motivation, teacher expectations, and of course a whole host of system-level variables. In complex environments these kinds of information can be as or more important in understanding interpersonal and institutional interactions than standard indicators on student achievement and teacher practice.

Secondly, even for standard measures, important information might also be only partially collected, or not systematically (for example, reasons underlying student drop-out or issues with teacher retention). It is true that the move towards more computerised environments and more accountable stakeholders has yielded a significant production of information that is now recorded by the majority of OECD countries (e.g. student achievement data, teacher assessment data, school budget data, school choices of parents etc.). However, the involvement of many stakeholders requires that the collected information is systematised in formats that make it easier to disseminate, analyse and compare nationally and internationally, and this is not always the case.

In addition, sharing information also comes with its own challenges. There might be few incentives for collected data to be shared widely, especially if there is a concern that it could be used in a negative manner (for example, in systems where there is strong competition for students between schools, the weaknesses of a particular school might be disguised or otherwise presented to avoid injuring the reputation of the school). Moreover, information might be hard to find, little publicised, or produced without thinking that it may be useful and hence not passed onwards to other actors in the system. The 2011 report of the Swedish National Agency for Education illustrates these problems; municipalities were shown to focus their attention on their rankings rather than performance, and used only a small share of the available data in its decision-making process (Swedish National Agency for Education [*Skolverket*], 2011).

On the other hand, too much information can obscure information pertinent to decision-making and/or render it unusable by its sheer magnitude. Loeb and Plank (2008) illustrate this danger with the California Education Code, which includes more than 100,000 articles and more than 2,000 pages. The abundance of information increases the difficulty with which stakeholders can learn about the existence of documents and then locate them. As O'Day (2002) points out, the abundance of information may even be counterproductive, as "teachers and schools may metaphorically and literally close the door on new information, shutting out the noise". It also raises the question of how all the information can be gathered and maintained in a way that can be used by the other parties. In a complex environment with multiple active stakeholders, harnessing all this heterogeneous data and ensuring that it reaches those who need it becomes an important challenge.

There are two elephants in the room in this discussion. The first is that increased access to data (via the Internet and including media-friendly testing and assessment results which lend themselves so well to league tables and rankings) does not ensure that the quality of the information is consistently high. The Internet has effectively removed many of the established gatekeepers or quality controls that were traditionally put in place by research institutes and academic journals (Burns and Schuller, 2008). As they argue: "More information is available, yes, but is it good information? And is it presented

accurately and in an understandable fashion? Can the reader use it in a comprehensible and useful manner?" (p. 17).

The second elephant is that the *production* of data does not equate to its *use* (Fazekas and Burns, 2012). In all discussions of increasing the availability of data in order to increase transparency and accountability to a broader range of stakeholders, there is an underlying equity issue. In most countries upper middle-class and middle-class families (or parent(s) with higher education, higher professional positions and higher income) are the ones that are most aware of how to actively use the education system for their own interest and benefit (Taylor, 2009). They are also more likely to have the capacity to lobby and press for change in the educational system through policy and practice (van Zanten, 2003).

In practice this means that in many OECD countries middle-class parents are more likely than parents with lower socio-economic status to use school achievement and school performance data, when available, in order to place their child in the best-performing schools. If changing schools is not possible, middle and upper-class parents are more likely to demand (and successfully lobby for) change in the system. Parents with lower incomes (including, in many countries, high proportions of immigrant parents) are less likely to be aware of their rights regarding school choice and may often lack the capacity to use achievement and performance data. A similar argument can be made that some districts or municipalities might be more likely than others to fully use available data – perhaps those that care more about education quality, or those that have better capacity to analyse and interpret such data. These equity arguments are not trivial – indeed, any system motivated to provide full access to performance and achievement data in the name of transparency and efficiency cannot turn a blind eye to how and by whom those data are being used.

The availability of data per se then, is not a stand-alone solution to information asymmetries between stakeholders, and can in fact serve to increase the complexity involved in their interactions. PISA 2012 analyses make it clear that simply making school achievement data public is not correlated with better student outcomes. Indeed, among the top performers in PISA 2012, Shanghai, China (number 1 in math achievement) is at the very bottom of the Figure 8.2, and Singapore (number 2 in achievement) is just slightly above the OECD average in terms public availability of school data. Information can only lead to school improvement if it is relevant, available in adequate quantity, and properly interpreted (O'Day, 2002).

Experimentation in a complex environment

The governance challenges outlined above have led to the rise of strategies to deal with complexity and the dynamic nature of the system, its actors, and the data available to them. Policy experimentation aims to improve the system by explicitly testing new policy options and assessing which could be successfully generalised. In addition to education, policy experiments are used in a number of domains, such as development aid, healthcare, economic policy, etc. They rely on small-scale trials, and have as their objective not the immediate improvement of the system, but rather the discovery of what key factors would generate such improvements. In complex environments, decision processes must typically reconcile conflicting objectives among stakeholders. It is here that experimentation provides strong arguments which can provide solid evidence on which to constructively base the discussion. However, it is precisely in complex environments that such a task is made more difficult, for reasons detailed below.

What is policy experimentation?

There exists an abundant literature on policy experimentation, but for the purpose of this chapter, we rely on Heilmann's (2008b) definition of experimentation as "a purposeful and coordinated activity geared to producing novel policy options that are injected into official policymaking and then replicated on a larger scale". This definition entails three key components of experimentation, which will be discussed in turn below.

The first important element is that experimentation is seen as *a deliberate process*, and must therefore be distinguished from innovations emerging by chance or as unintended side effects to other deliberate processes. Experimentation presupposes that a problem has been identified, has been understood as relevant, and that one or more potential solutions have been suggested that need to be assessed. It therefore offers a systematic approach to dealing with a problem: 1) identifying the problem; 2) suggesting a potential solution, 3) trying that solution out, and 4) evaluating whether this solution was effective. The deliberateness of experimentation resides in the planning of the process, and in particular, in the ex-ante decisions of what is going to be implemented and how its effects will be measured. Experimentation therefore entails a certain methodological commitment, which makes it different both from spontaneous innovation and from reactive policy making.

The second important element is that experiments are *performed at small scale* that can then be fully rolled out if proven to be successful. Underpinning this attempt is the objective of assessing whether the experiment could suitably be generalised to a wider implementation. This implies that it is evaluated and the results of that evaluation are used in the decision of whether to scale up or not. Experimentation thus entails a process in which evaluation is an explicit step, by comparing outcomes between areas where the experiment took place and areas where it did not, in an attempt to assess its efficacy. In order to be able to evaluate its impact and effectiveness, it is necessary for the experiment to include a carefully designed and chosen control or comparison group that can act as a reference point against which to measure the results of the experiment (see Box 8.1). In that respect, experimentation is an approach that greatly differs from whole-system reform, which approaches change from a reform point of view at the scale of the whole system. This will be discussed more fully below.

Box 8.1. Treatment and control groups in evaluation

To correctly assess the effectiveness of a policy or a new practice, it is not sufficient to measure certain outcomes following the introduction such policy; it is also necessary to assess what proportion of these outcomes can be explained by the policy itself rather than confounding factors. If a new reading method is introduced that increases reading scores by 5% among students, it is important to understand whether, for instance, reading scores were not just improving among all students that year. If that were the case, the 5% improvement might give the false impression that the method was effective, whereas it was actually due to other factors that contributed to a general improvement in reading scores.

To isolate the effect of a new method or policy itself from other confounding factors, it is necessary to have a counterfactual, that is, to find out what would have happened if that method had not been introduced. The best way to construct such counterfactual is to have a treatment group that is subjected to the new policy or method and a control group that is not.

Box 8.1. Treatment and control groups in evaluation *(cont.)*

To make sure the comparison does not capture other effects, it is important the treatment and control groups are as similar as possible in all dimensions, except whether they were subjected to the new policy or not. Difference in outcomes between treatment and control groups can then be attributed solely to the new method or policy being evaluated. Assigning subjects (students, schools, districts) randomly between treatment and control groups can help ensure that groups do not differ in any systematic way.

The third element of this definition is that this exercise in assessment is used as *feedback towards policy making*. Experimentation is by essence supposed to produce new knowledge. First and foremost it reveals the potential effectiveness of the policy being tried out. It also reveals additional information, for instance regarding possible unintended consequences that were not part of the initial thinking on that policy, information regarding the feasibility of generalising the policy to a wider scale, how well the policy is received among the stakeholders, etc. As such, experimentation represents a form of evidence-based policy making, in the sense that it generates new information that can be used by policy makers to make educated decisions about the direction in which to steer education systems.

Policy experimentation is not inherently linked to any particular methodology, and, as long as the previous conditions are met, can be argued to encompass both the strict standards of randomised controlled trials (RCTs) [2] and the qualitative case studies that mark much of education research. Pilot programmes probably represent the most widely used type of policy experimentation today, although it must be highlighted that they qualify as experimentation if (and only if) they contain a comparison group and a proper evaluation phase to assess the results.

However, part of the problem with pilot programmes as experiments is that researchers or policy-makers have often chosen the participants and/or location of the programmes, usually for practical reasons. For example, researchers interested in the effectiveness of a new writing software might very well choose to pilot it in a set of schools nearby, both because that would facilitate their access to the schools for testing and because they are more likely to have personal connections to the headmasters and school boards who approve research proposals and agree to participate in them. Similarly, a new curriculum with an emphasis on sports activities might be tried out in a small number of schools facing student behaviour problems; with the expectation that this where such a programme is more likely to work. But that reasoning creates a problem down the line: even if the programmes prove successful, it is hard to know whether the intervention would also be beneficial in other schools, and whether it should be generalised.

Randomised controlled trials (RCTs) try to address this issue by assigning participants (students, classes, schools, teachers, etc.) randomly between a treatment group (where the new policy will be implemented) and a control group where conditions stay the same as before the experiment (see Box 8.1). By constructing two groups that are on average identical in their composition, any difference in the measured outcomes between the treatment and the control group can then likely be attributed to "treatment", i.e. the policy being tried out.[3]

RCTs therefore offer the cleanest way to estimate the effect of a given intervention while being able to exclude confounding factors[4]. RCTs also allow for manipulating several variables at once, by setting up an experiment with several treatment groups in which experimenters vary some parameters, e.g. the length of the treatment, how many hours a week are devoted to it (intensity), etc. For instance, when Schwerdt et al. (2011) tested for the effectiveness of adult learning vouchers, they randomised the monetary value of vouchers, and could check whether their effectiveness depended on their amount (it did not). By being able to test for multiple treatments at once, RCTs are in this sense economical compared to pilot programmes which, by essence, can only test for one set of parameters.

But such clarity and testing power comes at a price: RCTs typically require samples (of students, classes, schools, etc.) big enough that it is possible to attribute difference of outcomes between treatment and control to the intervention rather than to chance. In addition, while they may tell you what works (or does not work), they do not generally answer the question of *how* and *why* something works. Understanding those elements requires additional evidence often gleaned through other methodologies (OECD, 2007).

Sometimes, a change or reform occurs outside the scope of a carefully designed experiment. While prior design is always preferable from an inference point of view, it might in some cases still be possible to use quasi-experimental methods to estimate *ex-post* the impact of the change. In particular, when there are reasons to believe that the change affected subjects at random[5], the same comparison analysis as in RCTs is still valid.

In the field of education, regression discontinuity design (RDD) is often used to estimate the effect of certain factors in quasi-experimental settings, and is even recommended when random assignment is not possible. The technique compares subjects on one side and the other of a boundary that determines to what policy they are subjected (the boundary creates a discontinuity in policy). The boundary can be a physical or geographical boundary (e.g. district limits), a threshold grade, the birth month or year, etc. For instance, van der Klaauw (2002) compares students with scores closely under and above the threshold to qualify for financial aid, in order to estimate the returns to college attendance. The idea of RDD is to take a narrow band of subjects around the threshold, so that the groups are similar and can be treated as if they had been randomly assigned above or below. In general, these quasi-experimental approaches are harder to control, but research among education stakeholders using these designs can be encouraged in order to gather helpful information without necessarily having to incur heavy costs that can sometimes affect pilot programmes or randomised experiments.

Although experimentation requires the collection and analysis of quantitative data to evaluate the treatment being considered, it does not exclude other more qualitative tools. Experimentation does not occur in a bubble. Qualitative research such as surveys and case studies can yield important insights in a number of ways. Qualitative techniques can be used to help select hypotheses to be tested and refine the design of the experiment. After and during an experiment, qualitative techniques can help tease out various mechanisms and uncover potential explanatory factors and secondary effects.

Inference in a complex environment

In order to improve systems, researchers have argued that changes should be implemented in a way that could foster learning opportunities about what works and what does not (Campbell, 1969). The ability to learn from a particular experiment is inherently

linked to the quality of its original design. However designing and conducting research that yields this information is not always easy: in particular, a randomised sample might be difficult to obtain in many fields, including education (Campbell and Stanley, 1963). This is true for a number of reasons, with the most pertinent one being that participation in educational experiments might be hard to enforce. In such a situation, attrition could lead to a smaller sample, a biased sample[6], or simply to the experiment being cancelled (see Box 8.2).

Box 8.2. The problem of attrition in education experimentation

Attrition refers to the fact that some entities (schools, classrooms, students) may drop out before the experiment is scheduled to finish. This can be problematic because it is harder to assess differences between treatment and control groups once the sample size becomes too small. In addition, there is a possibility that attrition is not random, but rather biased in a way that would affect the comparability of the treatment and control groups. For instance, perhaps wealthier parents would choose to relocate their children to non-participating schools, thereby leaving on average poorer students in the one group. This can then give a false picture of treatment effectiveness. Although it is possible to compare the composition of both groups after the fact, selective attrition cannot be corrected for in statistical analyses.

In education, this can be particularly problematic because participation in experiments is often voluntary. Since the intervention is new, stakeholders might not always be willing to participate. Offering incentives (monetary or otherwise) to participate in the experiment has been suggested as one way to reduce attrition or outright refusal to participate in an experiment. The downside is that it increases the cost of the experiment. It might also serve to select participants that are more responsive to such incentives, and thus not be representative of the entire population.[7]

Attrition raises a key issue for experimenters as *inference*, the task of learning from the data generated by experiments, can only be done when it is possible to compare the outcome of the experiment with a reasonable counterfactual. In a complex environment, inference of cause and effect becomes, well, complex. Even randomised controlled trials (RCTs) provide estimates for what would be the average effect of a certain policy, holding everything else constant. They identify what would happen on average if such a reform were to be extended to a similar group. However, the notion of an average effect might not be relevant in an environment that is characterised by important idiosyncrasies and non-linear dynamics. For instance, an experiment involving the provision of free textbooks for students might have very different effects depending on whether it is applied in a wealthy district or in a district where households face harsher budget restrictions. Effects can also be non-linear when a policy requires a critical size to be effective, for example with vaccination campaigns, which are ineffective unless they reach a significant share of the population. It may be important to understand how a potential policy might affect various contexts differently and if possible to find potential explanations for such variation.

Another complicating issue for experimentation, particularly salient in the realm of education, is that individuals may exhibit different behaviours if they know they are part of an experiment – a phenomenon sometimes called observer effect or Hawthorne effect. When this takes place, the whole subject pool (both control and treatment group) becomes systematically different from the rest of the population.[8] Left ignored, such observer effect might give a distorted picture of what would happen if the experiment

were to be generalised. In practice, this effect is more likely to take place in situations where individuals are expected to provide some efforts towards a task. For instance, in their experiment on low-stake feedback to teachers in Andhra Pradesh (India), Muralidharan and Sundararaman (2010) found that teachers altered their classroom behaviour even when they were assigned to the control group. While it is theoretically possible to remove this observer effect by comparing with individuals who are not part of the experiment, it is not always feasible to do so, for practical or ethical reasons.

Lastly, inferring the effect of a policy is accomplished by isolating it from confounding changes that occur at the same time. This *ceteris paribus* assumption, i.e. that all other things (than the policy) are held constant, might not be a trivial one to make in a complex environment, characterised as it is by fluctuating contexts and tipping points. The existence of feedback loops, especially reinforcing ones (usually termed positive feedback), may generate systems that exhibit several equilibria between which a system might oscillate, or on the contrary be permanently stuck. The capacity of the system to absorb changes, or on the contrary to exacerbate them, could then generate very different responses to the same input. When such path dependence[9] exists, a given reform might yield very different outcomes even when applied to seemingly similar systems. Recent research in the Netherlands (van Twist et al., 2013) has demonstrated that for example the assignment of the label "very weak" to a school can elicit a positive response from one school and a negative response from another, depending on the local context, history and staffing situation at the school; in self-reinforcing processes. Such sensitivity renders the analysis of the causes of success or failures for any particular policy intervention more complicated.

Bottom up experimentation is important but not sufficient

The salience of local discrepancies between schools or districts may suggest that bottom up initiatives are the best way to deal with educational challenges. If schools vary greatly from one another, then it is possible that solutions emerging from each school would be better suited to tackle their respective problems. However, in a complex system, bottom up initiatives cannot be scaled up to the broader system without at least some level of centralised discussion. As such in such a context, experimentation offers a systematic approach to this discussion.

The Dutch example outlined above (van Twist et al., 2013) illustrates a key insight. Namely, when a system exhibits self-reinforcing dynamics[10], some elements will benefit from virtuous cycles (success breeds success), while others will be caught in vicious circles (where difficulties bring about further difficulties). The consequence of such differences in local dynamics means that some heterogeneity will exist and persist in the system. Small initial differences in local contexts can therefore be exacerbated[11], creating a situation in which important discrepancies between schools or districts can persist and become hard to mitigate. Drawing from the Dutch GCES case study (van Twist et al., 2013), after receiving the "very weak" rating, some schools were facing a vicious circle of attrition among their students and staff, triggered by negative assessments by the Inspectorate; other schools experienced a virtuous cycle where parents would trust the school to perform adequately, which in turn led to a cooperative culture in which school staff and parents could work together. These discrepancies imply that careful attention must be paid to the particularities of each educational context.

The importance of context in education may suggest that local solutions might be a useful way to identify new, broader, policy options. However, scaling up local solutions

is hampered by the specificity of particular success stories. Generalising the success of one initiative to another school, district or wider area may generate very different results from the ones initially expected. As a result, although bottom up initiatives are an important way to generate innovative ideas from the field and suggest new policy options, decisions about rolling them out on a larger scale must be taken within the context of a centralised discussion about their potential for generalisation and sustainability on a larger scale.

Heilmann's analysis of policy experimentation in China can shed an interesting light on the experimentation process in a geographically large and culturally diverse context. He argues that China's economic success can be partially explained by the so-called "point-to-surface technique", an approach to experimentation which "gives room to local officials to develop models on their own, while ultimate control over confirming, revising, terminating and spreading model experiments rests with top-level decision-makers" (Heilmann, 2008a). The approach lets individual decision-makers (points) experiment and implement new ideas, which are then reported higher-up in the hierarchy, and if approved, then generalised to the rest of the constituency or country (surface).

This freedom in experimentation at the local level, later sanctioned or discarded by higher levels of government to fit within the central government's objectives, allows responding to local specificities of each economic area. However, Heilmann's analysis highlights the importance of having information and results from (possibly spontaneous) local initiatives and experiments disseminated to higher levels of decision making, which then decide whether to implement the experiment on a wider scale. Although the specificities of China's governance process do not map well on many OECD countries, it is an intriguing example of one approach to addressing what has been labelled the "tight but loose" issue (Thompson and Wiliam, 2008, cited in Hopfenbeck et al., 2013). That is, an approach to implementing reform (or in this case system level change) that keeps in mind core central principles (tight) while leaving room to implement these principles according to an individual's or region's initiative (loose).

It is a similar intent that was behind the change made in 2005 by Flanders in Belgium, which set up a three-year period during which schools were allowed try out temporary projects and experiments in pre-specified list of domains. As a result, individual schools applied for authorisation in order to experiment with things such as curriculum, extra help to non-native speakers, teacher autonomy, contractual arrangements, peer-review of teachers, etc. The idea was to allow schools to experiment, and later on take stock of what was successful or not, with a view to feed those insights into policy-making. That said, it is debatable whether such a non-systematic approach could yield learning robust to other contexts. The evaluation report was limited when drawing general conclusions from those varied experiments (Ministerie van de Vlaamse Gemeenschap, 2009).

Indeed, despite the intentional element to the inductive approach to experimentation, illustrated by the Chinese and Flemish examples, important subjective judgements are still made as to what could be successfully scaled up. As a result, while using unfettered local experimentation can stimulate the emergence of new ideas, this type of analysis still relies on post-hoc selection of initiatives based on the subjective assessment that they are transferrable and that they comply with grander (often politically motivated) objectives.

The use of carefully designed policy experimentation can offer a more harmonised framework in which policies can be tried, at a smaller scale, and in various contexts, as it can substitute for a more systematic approach to finding and scaling up novel policy solutions. In paying attention to the way the experimentation is deployed, it can also help

identify what contextual factors can affect success or failure, an important element to take into account when determining whether, where, and how the experiment can be successfully scaled up. Experimentation therefore attempts to remove the subjective part of the judgment, and replace it instead with a more systematic approach that identifies success criteria ex-ante.

Experimentation in the debate between whole-system change and gradual reforms

In complex environments, all agents are embedded in a network in which they interact with each other. This means that ultimately, the success or failure of certain educational practices can be only assessed by taking all stakeholders into account. There are two contrasting ways to approach this challenge. One approach emphasises a whole-system approach to change in order to tackle concomitant issues at once; the other approach emphasises instead the necessity of small changes in order to more clearly identify the effects of the intervention and bring stakeholders on board. Experimentation falls into that later category. What follows is an overview of the two perspectives.

The whole-system approach argues that change should occur at a greater scale in order to capture all relevant nodes of a system. Targeting change too narrowly might be unsuccessful, because it affects only a small part entangled in a much bigger network. This whole-system approach is for example suggested by Fullan (2009), who welcomed the emergence of numerous intentional system-level reforms since the mid-1990s and advocated a multilevel approach to change that encompasses schools, regional subdivisions/districts and the central government. This is also the approach suggested by Mason (2008), who argues that feedback loops endow complex systems with a significant "inertial momentum". Consequently, "what it might take to change a school's inertial momentum from an ethos of failure is a massive and sustained intervention at every possible level until the phenomenon of learning excellence emerges from this new set of interactions among these new factors, and sustains itself autocatalytically[12]."

This echoes arguments made for instance by Lipton and Sachs (1990) about the transition of Eastern Europe to market economies in the 1990s. They argued that states should take advantage of their *état de grâce* to implement a "big bang" approach to (economic) reforms, sometimes referred to as "shock therapy". In the realm of education, recent experience has for instance shown that the publication of PISA results has, for some countries, constituted a wake-up call that prompts public discussion on education and yields momentum for sweeping reform.

While a whole-system approach might be a successful solution to facilitate change in a complex environment, it might be problematic from a learning point of view. The goal of experimentation is fundamentally that of learning: experiments are designed to assess the effectiveness of new policies or practices in order to offer new viable policy options. The whole-system approach to reform runs counter that objective by offering a blanket approach to change.

Popper (1957) argues against whole scale change on this epistemological ground and instead favours a "piecemeal" approach, sometimes referred to in the literature as "gradualism". Popper argues that a piecemeal approach "permits repeated experiments and continuous readjustments". Since learning occurs through trial and error, Popper emphasises the importance of being able to identify the causes of success or failure of a change. According to Popper, as the holistic approach does not offer a counterfactual, it is not possible to identify whether successes or failures are due to the policy, or whether

outside elements confound the results. It then becomes impossible to eliminate a bad policy and replace it with something better. Instead, he favours a more gradual approach that operates on the basis of successive incremental changes. It then becomes possible to assess success or failure to each of these steps, and to have a clearer picture of what works and what does not. This is also the point made by Campbell and Stanley (1963), who argues that, as much as possible, reforms should be rolled out in a way that can lead to proper evaluation. For instance, when random assignment is not an option, they offer staggered interventions as a possible alternative to be able to compare groups in a meaningful way.

A further argument in favour of gradualism and experimentation can be made on more pragmatic grounds: gradual changes are easier to accept and therefore to implement. This is in essence the point made by Dewatripont and Roland (1995), for two reasons. First, gradualism can help build support for reforms over time: initial successes can form a basis on which to build popular support for subsequent changes. Second, gradualism ensures that if the initial reforms do not work as intended, the cost of reversal is lower than if a larger collection of reforms had been implemented.

These two opposed views highlight an implicit tension in complex systems, between learning and realism. On the one hand, understanding whether a policy would yield the desired effects typically requires a smaller-scale well-designed experiment, which allows for rigorous comparison with (near) counterfactual settings. This enables a more precise evaluation of the policy option being considered, the effectiveness of which can therefore be assessed in order to determine its suitability for generalisation. On the other hand, when operating on the basis of narrow interventions one risks missing the effects of certain reforms that require several changes to be made simultaneously or that require a critical size to be effective. In such a case, the interconnectedness of the system constitutes a threat to the external validity of the experiment: the change might have different effects if it is scaled up. One must then adopt a more holistic approach that takes that interconnectedness into account if one aims to understand the effects a policy would actually have if once scaled up. And of course, there is always the question of what is feasible given existing political and budgetary constraints.

This tension is echoed by similar debates in other fields. For instance researchers and policy makers working on developing countries have debated whether economic interventions should take a big-package approach, in which multiple objectives are pursued at one time (see for instance the Millennium Villages Project[13]), or a more targeted approach that enables more learning (Banerjee and Duflo, 2011). This debate in economic policy parallels the debate between whole-system approaches versus piecemeal changes, in that whole-system tries to generate a big change momentum by addressing many problems at once, but presupposes that the solutions they offer are the correct ones, whereas targeted change takes a more cautious and agnostic approach, but can only work on smaller well-identified problems.

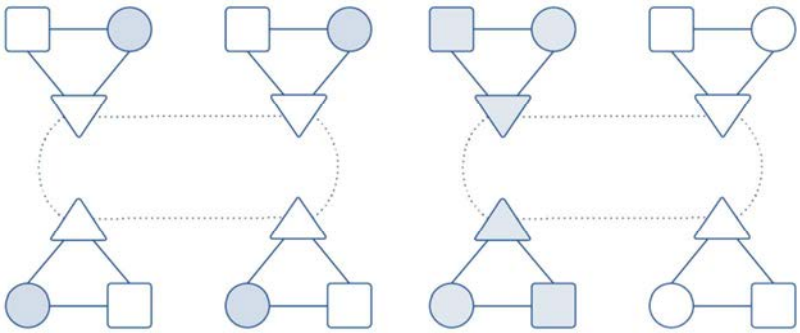
That tension between the possibility of experimentation and the accuracy with which it can be done (taking complexity and interconnectedness into account) can never be fully resolved. Since the goal of experimentation is to learn, experimenters and designers ought to err on the side of narrow but well-identified experiments. However, a good balance can sometimes be struck by designing experiments that operate on units of significant scales, or ecosystems. This is detailed below.

Moving towards ecosystem experimentation

The challenge of complexity is acknowledging that while no perfect solution exists, it is possible to take small concrete steps to make a difference. Policy experimentation can be made more useful by choosing an appropriate scale and design, called *ecosystem experimentation*; by adopting the policy cycle to reflect the dynamic nature and the intricacy of education systems; and by ensuring that input from stakeholders is also matched by a culture of constructive criticism within the system that can identify successes and failures.

Embracing ecosystem experimentation involves moving from horizontal experimentation, where experiments focus on a certain type of node (e.g. changing the reading teaching method in all schools), towards experiments are conceived as focussing on self-contained parts of the systems (i.e. natural ecosystems).

Figure 8.3. Horizontal experimentation (left) versus Ecosystem experimentation (right)



Note: Dotted lines denote weak links; solid lines denote strong links between elements of the system.

Figure 8.3 illustrates the distinction between those two concepts. On the left panel, the horizontal experiment targets all nodes of a certain type (in this case, the circles in grey, which may for example represent schools[14]). On the right panel, the experiment focuses on self-contained subsets of the system, for instance the networks of all stakeholders within a small set of given districts. The idea behind targeting self-contained subsets is to identify pockets of the system in which the links and influences within the subsystem are much stronger than links with other parts of the system. In the figure above, these pockets are represented by a triad circle-square-triangle. Note that the interactions are strong within a triad (solid lines), while interactions *between* triads are weak (dotted lines).

However, the reality of today's education systems is that no part of the system is completely isolated from the others. Instead, many outcomes (in the broad sense) result from the interactions of many actors. Let's consider the example of healthy dieting, which has been identified as a crucial element of education (OECD, 2014). Concerns about increasing obesity have emphasised that students' eating habits must involve parents, schools and teachers. From an education perspective, an experiment trying to change students' eating habits might try a different layout of cafeterias in a number of randomly selected schools, a horizontal experiment. But parents constitute an important factor in children's diet, and only changing food available in school might not suffice. Instead, school menus and parental food choices might reinforce each other (for better or worse).

When such complementarities exist, it might be more fruitful to randomize at a higher level (e.g. assign whole towns to either treatment or control) and involve both cafeterias *and* parents in the experiment (e.g. by providing parents with information, workshops, etc.). Such an ecosystem experiment would be able to capture the joint effect of school food and parental influence (without, unfortunately, being able to distinguish their respective effect).

When trying to target ecosystems, a new question emerges: where to draw the boundaries of any given ecosystem? In an intricate environment it might seem *prima facie* difficult to isolate subsystems that are self-contained. However, research in ecology shows that in complex networks, each node is mostly affected by other nodes within two or three degrees of separation (see for instance Williams et al., 2002). Applied to the educational paradigm, this means that it is often possible to isolate a relatively separate subpart of the system, in which nodes strongly influence each other but are not too affected by the other part of the networks[15]. In a country where education is managed at the district (or municipality) level, these could represent independent subparts of the system. In metropolitan areas, neighbourhoods might constitute such subparts. The definition and boundaries of such ecosystems will of course depend on the nature of outcome considered. One example of this multi-stakeholder experiment is the work of the Harlem Children's Zone (HCZ), a wide-ranging programme set in Harlem, New York.[16] From its inception, the objective of the HCZ programme has been to tackle a number of issues simultaneously, both inside and outside schools, in order to create a tipping point towards a positive learning environment. For eligible communities, the intervention combines two elements. First, children can benefit from a pipeline of charter schools, from parenting classes, followed by early childhood interventions, all the way to high schools, and support for college entry. The schools in the programme also offer extra-hours, social activities, healthy foods, sport programmes, as well as tutelage in non-cognitive and employability skills. In parallel to that pipeline, the second tool of the HCZ programme is to invest in community programmes, such as cooperatives, health and food education to families, healthcare initiatives, community centres and foster care prevention. The programme therefore combines a move towards charter schools, with investments in the surrounding communities, in order to involve a wide range of education stakeholders. The HCZ has been estimated to be very successful at bridging the racial achievement gap in the targeted area (Dobbie and Fryer, 2011.

Perhaps the most important consideration when choosing the scale at which the experiment takes place, and the level at which to randomise, is that the best scale often depends on the relevant type of interaction between agents being studied, and therefore on the question being analysed. In education, one of the most common concerns is therefore whether to randomise at the student level, at the teacher level, at the classroom level or at the school level.[17] Some interventions might even want to consider randomisation at the school district level. In general the statistical analysis is made easier if the randomization takes place as close as possible to the final level of analysis. So, when investigating the effect of a given intervention on student's scores, it would seem easier to allocate students randomly between treatment and control groups, for example.

However, in certain experiments where there are significant interactions between students, student-level randomisation might not be appropriate. Consider for instance a hypothetical experiment that would look at the effect of providing more information about future careers on students' motivation. Since peer effects could be important factors that affect educational and professional aspirations, it might make sense to account for peer connections between students when designing the experiment. Randomising at the

student level would lead to classes comprised of students from both the treatment and control groups. If students within classes interact with each other extensively, it is likely that information provided to students of the treatment group would leak to students of the control group; or that behaviours from one group could affect behaviours of the other group's members. This "contamination" would likely make outcomes of treatment and control groups more similar; as a result, it may give experimenters the – possibly wrong – impression that the experiment did not have any effect. This is a situation where it might make more sense to allocate whole classrooms, or even whole schools between treatment and control, in order to minimise the spread of information, and other contaminations effects, between treatment and control. In general, experiments that involve information campaigns are more likely to suffer from issues of contamination, and are probably better randomised at some reasonably high level, such as school- or district-level. It is however useful to remember that this comes at the cost of reduced statistical power.

As mentioned, another question that might arise during the design of certain education experiments is whether to randomize at the teacher or at the classroom level.[18] Consider the example of a new teaching method, say in mathematics. To assess its efficacy, it might useful to train some teachers so that they use that new method in class, while others stick with the traditional method. But doing so creates the risk, although improbable, that the teachers selected are not representative of the other teachers; if this were the case, then assessing the efficacy of the new method based on those teachers would give a wrong picture of what would happen if the method were extended to other teachers. A potential solution to remove the "teacher effect" is ask trained teachers to use only in some of their classes (chosen at random), and compare outcomes in classrooms where the technique is used and classrooms where it is not. For this setup to be valid, it requires that teachers can "turn off" the new method, and teach in two different ways, without the new method contaminating the way they teach traditionally. If the experiment were instead an information campaign towards teachers, e.g. reminding them about gender issues among students, this might be problematic, as it is difficult to imagine that teachers could "turn off" this newly-acquired information. In that case, classroom-level randomization would likely suffer from contamination through the fact that same teachers teach different classes. It would in such case be more appropriate to randomize at the teacher-level. Again, the scale to choose for the experiment, and in essence, the appropriate definition of what constitute an ecosystem, depends on the question being examined.

Paradoxically, while complex systems might be where ecosystem experimentation is the most necessary, it will also be where they are the hardest to implement in practice. Since randomised designs aim at obtaining a precise evaluation of an experiment, they are particularly vulnerable to attrition. What if the units selected for the policy trial would rather not participate? In many countries, ministries might not have the power to enforce enrolment in the program, nor might they be able to prevent opting out while the process is underway (see Box 8.2 on attrition above). For such a design to function, it is imperative that excellent communication and promotion to all stakeholders involved are present. Such communication must clearly explain the content of the approach, how everyone will be affected, and the question the experiment is trying to answer. A key benefit of involving all stakeholders at once is that it avoids the pitfall of ex-ante singling out one stakeholder as a bottleneck towards educational achievement.

It is also important for all stakeholders to understand that their participation is crucial for the experiment, regardless of the outcome. Although an experiment could reveal an amazing improvement in the variable under study, it could also reveal that there is little,

or no, effect of the intervention. Although not as "sexy" as observing strong improvement, knowing when something does not work – or works but only to a limited extent – is also crucial to making policy[19]. In times of limited resources, it is extremely useful to know what will give the greatest return on investment, and this information can only be learned through trying things out – ideally, through experimentation.

Adapting the policy cycle

How would this work in practice? The premise of the complexity approach to education is that success or failure of schools depends on many interlinked stakeholders who operate in a fluctuating environment. Only by involving all relevant stakeholders at once is it possible to generate a switch from one type of equilibrium to another, in order to overcome what Mason (2008) calls "inertial momentum". A key tool for ensuring the involvement and participation of all stakeholders is to move towards a more "open source" input approach to experimentation process. By widening inputs for ideas and comments, notably through more integrated IT systems (Loeb and Plank, 2008), it is possible to make sure that all stakeholders can participate in the identification of problems, have their say in their prioritisation and can suggest potential solutions (Box 8.3).

Box 8.3. Technologies in education governance

Information technologies have affected the way teaching takes place, and also the way schools, teachers, students, administrators and local authorities interact with each other. In addition to using information and communication technologies (ICT) to manage scheduling, staff, students and grades internally, many schools have also turned to ICTs to facilitate contact with parents and the wider community and promote more inclusive governance and accountability. For instance, the Eudora school district in Kansas (USA) has set up explicit policies and guidelines to facilitate online presence by teachers, school staff, and teacher-parent contacts, to both share everyday activities and to inform and deal with problematic situations (Lepi, 2013).

In open data initiatives, cities or governments decide to make their data freely available. The data is designed to be accessible to local authorities, schools and headmasters, and parent and community members. Relatively recent, these initiatives are becoming widespread: examples in education include the MySchools website in Australia, the New York City Department of Education's School Choice Design Challenge, Sweden's Open Comparisons website, and SIMCE in Chile.

Frequently inspired by issues of school choice and tracking student achievement, these trends highlight the greater involvement of parents and the community in the governance of education systems. This involvement can be formally structured and solicited, as in the examples above. But it can also be informal and rely on the power of ICTs and social media to unite actors brought together by a shared willingness to improve education in their communities.

Policy experiments can provide essential evidence when considering whether an evaluated policy should be rolled out on a greater scale. In particular, if the experiment is properly designed, its results should be used to inform future policy making and feed reflexion on the causes of success or failures. Experiments can and should constitute important elements to hold policy makers accountable: they should be able to justify costly wide-scale reforms using sound knowledge, including experiment-based knowledge. The analysis and dissemination of the results are not designed to blame

specific individuals but rather as necessary feedback for a better understanding of what works, what does not, and how this new knowledge can be taken further. It is a necessary condition for establishing a trial-and-error system of successive experiments in which constant learning is possible. The knowledge for all participants that they are ultimately contributing – possibly in an indirect and implicit way – to the improvement of future generations' education systems and practices, can be a strong motivating factor.

The process of experimentation should also be able to adapt to changing conditions, by continuously adjusting its methods and processes. As such it may be possible to design experiments that undergo several iterations to address successive as a form of "experiment package" designed to address related questions at once (Cobb et al., 2003) or to successively narrow down the factors of interest that may help explain specific outcome dimensions (Box, 1999). However, this must be done with caution. A good balance needs to be achieved between the ongoing tweaking required to adjust to feedback loops and the need for experiments to be sustained for long enough that their effects have time to set in. Minor changes can still be accommodated reasonably easily in the experimentation design without nullifying all previous data, but attention should be devoted to ensure that experiments are carried on long enough for meaningful data to be gathered.

Conclusion

This chapter analysed the impact of complexity on policy experimentation in educational systems and suggested potential ways that it can be harnessed. The difficulty of experimentation is that it requires the involvement of many interdependent stakeholders with different cultures, motivations and time horizons, in a careful exercise of co-ordination towards a common goal of identifying policies that work and policies that do not. Governing in such an environment poses many challenges.

The complexity approach acknowledges the interdependence of all stakeholders in education systems, and sees managing this interdependence as a key factor in educational achievement. In such a context, a linear approach to policy-making will not work adequately. The process of identifying issues and suggesting novel policy options to solve them must therefore rely on experimentation, which can be a source of solutions that can take account of local contexts while preserving a systematic approach to solving issues.

When turning an idea into an actual experiment, designing the implementation of the experiment adequately must take this complexity into account. Experiments must strike an appropriate balance between two antagonistic objectives: on the one hand, involving all the stakeholders that could affect the experiment's success or failure, which requires operating at a wide enough scale; on the other hand, designing an experiment that can be rigorously evaluated, which requires experimenting with targeted changes. Although these objectives can never be fully brought together, the chapter suggests *ecosystem experimentation* as a potential solution: identifying self-contained parts of educational systems, such as districts, and randomising their allocation between treatment and control.

Notes

1. For a more detailed analysis of complexity and its impact on educational governance and reform, see Snyder (2013).
2. Such experiments involve a wider sample, divided at random between those who are allocated to the treatment (the policy being tried out), and those who serve as a control group. This design ensures maximum comparability of the groups in order to better assess the effects of the treatment. In the field of education, this was famously pioneered by experiments such as the Tennessee STAR (Student-Teacher Achievement Ratio) experiment in the United States, to assess the effects of class size on student learning, or more recently the extensive work by Fryer (2011) on what types of incentives affect student outcomes.
3. It is possible to construct experiments that allow for the comparison of multiple policies (one at a time), by constructing as many treatment groups as there are of policies to be tried out, while retaining one control group. This of course requires a bigger pool of participants.
4. The crucial idea behind RCTs (and in general any statistical analysis trying to estimate the causal effect of one variable on another) is that the variation in the causal factor should be unrelated to other factors that could affect the outcome. If one allocates subjects randomly, by definition the causal factor (being in the treatment instead of the control group) will be unrelated to other characteristics that might affect the results (e.g. age, ability, socio-economic background, etc.). It then becomes possible to establish the causal effect of the treatment on the outcome (e.g. student literacy rate), while excluding possible confounding factors.
5. Researchers talk about a "natural experiment", as if nature had allocated at random subjects between treatment and control groups.
6. If attrition is correlated with certain individual characteristics, the sample will not be representative anymore.
7. Despite this selectivity problem, policy makers may sometimes be interested in the effect of *offering* a certain programme, where take-up is ultimately voluntary.
8. A good overview of problems of attrition and biased subject pools in field experiments can be found in Harrisson and List (2004).
9. Path dependence is the property of systems in which the final outcome depends extensively on previous conditions, so that a small change in previous conditions might yield large changes in outcomes. A tennis ball placed on the top of a pitched roof exhibits path-dependence because a small push in one direction or the other determines on which side of the house the ball will fall. A counter-example is a marble dropped in a convex bowl, which always ends up at the bottom of the bowl regardless of the initial starting point. Systems that exhibit positive reinforcement loops can easily exhibit path dependence (Page, 2006). One education example is when reputation can help achieve better performance. For instance, if parents can choose where to enrol their children, a school that has a better reputation can attract more students and is therefore able to select the highest achievers more easily. Conversely, a school that is expected to perform poorly will fail to attract good students, and as a result confirm the initial prediction, a self-fulfilling prophecy.

10. These self-reinforcing dynamics are sometimes referred to as positive feedback loops. They exist when the more a phenomenon takes place, the more it feeds itself.
11. This property is also known as path dependence.
12. Autocatalytic reactions are chemical reactions that produce the catalytic compound which enables the reaction to take place. This means that once they start taking place, they fuel themselves; Mason uses this metaphor to talk about dynamic changes in education systems that can sustain themselves once they have reached a critical threshold.
13. See www.millenniumvillages.org/ (accessed 16.12.2015); see also Sachs (2006).
14. Instead of certain entities, such as schools, experiments can also target processes, such as teaching methods, or reporting workflow in schools, etc. The argument remains identical: a consistent set of processes should be targeted at once to have a more comprehensive understanding of the total effect of the experiment. This does not preclude the use of randomization and control groups.
15. For examples of how this identification might work in practice, see Snyder (2013).
16. See Harlem Children's Zone project website, www.hcz.org
17. Randomising at the student level means allocating individual student randomly between treatment and control group. Randomising at the teacher level means that teachers are individually (and randomly) allocated between treatment and control; if their students are part of the experiment, then all students of a given teacher should be allocated to the same group (treatment or control) as that teacher. And so on for school- or district-level randomization.
18. This echoes concern in clinical medicine about whether to randomize at the patient level or at the practitioner's level.
19. Clearly this is a difficult message from a political point of view. Communication with the stakeholders should thus make clear that experimentation is in fact also about efficiency, in the sense that understanding the effectiveness of a suggested policy is good value for money, if even the result is neutral or negative.

References

Angrist, J.D. and V. Lavy (1999), "Using maimonides' rule to estimate the effect of class size on scholastic achievement", *The Quarterly Journal of Economics*, Vol. 114/2, pp. 533-575, http://dx.doi.org/10.1162/003355399556061.

Banerjee, A. and E. Duflo (2011), *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*, Public Affairs, New York.

Borman, G.D. et al. (2003), "Comprehensive school reform and achievement: A meta-analysis", Review of Educational Research, Vol. 73/2, pp. 125-230, www.rer.sagepub.com/cgi/content/abstract/73/2/125.

Box, G.E. (1999), "Statistics as a catalyst to learning by scientific method, part II – a discussion", *Journal of Quality Technology*, Vol. 31/1, pp. 16-29.

Brown, A.L. (1992), "Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings", *The Journal of the Learning Sciences*, Vol. 2/2, pp. 141-178.

Brown, L. and S.P. Osborne (2013), "Risk and innovation", *Public Management Review,* Vol. 15/ 2, pp. 186-208, http://dx.doi.org/10.1080/14719037.2012.707681.

Burns, T. and T. Schuller (2008), "Evidence-informed policy in education: New opportunities and challenges", in R. Desjardins and K. Rubensen (eds.), *Research of Policy and Research for Policy in an Era of Transnational Education Policy Making*, Aarhaus University Press, Aarhaus.

Byrne, D. (1998), *Complexity Theory and the Social Sciences: An Introduction*, Routledge, London.

Campbell, D.T. (1969), "Reforms as experiments", *American Psychologist*, Vol. 24/,4, pp. 409-429.

Campbell, D.T. and J.C. Stanley (1963), "Experimental and quasi-experimental for research on teaching", in N.L. Gage (ed.), *Handbook of Research on Teaching*, Rand Mc Nally & Co., Chicago.

Chingos, M. M. and G. J. Whitehurst (2011), *Class Size: What Research Says and What it Means for State Policy,* Brookings Institution, Brown Center on Education Policy, Washington, DC, www.brookings.edu/research/papers/2011/05/11-class-size-whitehurst-chingos.

Clay, E.J. and B. Schaffer (1984), *Room for Manoeuvre: An Exploration of Public Policy Planning in Agricultural and Rural Development,* Heinemann, London.

Cobb, P. et al. (2003), "Design experiments in educational research", *Educational Researcher*, Vol. 32/1, pp. 9-13.

Dewatripont, M. and G. Roland (1995), "The design of reform packages under uncertainty", *The American Economic Review*, Vol. 85/5, pp. 1207-1223, www.jstor.org/stable/2950984.

Dobbie, W. and R.G. Fryer (2011), "Are high quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone", *American Economic Journal*: *Applied Economics*, Vol. 3/3.

Fazekas, M. and T. Burns (2012), "Exploring the complex interaction between governance and knowledge in education", *OECD Education Working Papers*, No. 67, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k9flcx2l340-en.

Fryer, R.G. (2011), "Financial incentives and student achievement: Evidence from randomised trials", *The Quaterly Journal of Economics*, Vol. 126/4, pp. 1755-1798, http://dx.doi.org/doi:10.1093/qje/qjr045.

Fullan, M. (2009), "The large-scale reform comes of age", *Journal of Educational Change*, Vol. 1/1, pp. 5-28.

Grossmann, D. (2014), "Secret Google lab 'rewards staff for failure", webpage, BBC News, www.bbc.com/news/technology-25880738 (accessed 27 October 2015).

Hallsworth, M., S. Parker and J. Rutter (2011), *Policy Making in the Real World,* Institute for Government, United Kingdom.

Harlem Children's Zone, (n.d.), "From Harlem Children's Zone", webpage, www.hcz.org/?s=From%20Harlem%20Children's%20Zone (accessed 1 March 2014).

Harrisson, G.W. and J.A. List (2004), "Field experiments", *Journal of Economic Literature,* Vol. 42/4, pp. 1009–1055, http://dx.doi.org/10.1257/0022051043004577.

Heilmann, S. (2008a), "From local experiments to national policy: The origins of China's distinctive policy process", *The China Journal*, Vol. 59, pp. 1-30.

Heilmann, S. (2008b), "Policy experimentation in China's economic rise", *Studies in Comparative International Development,* Vol. 43/1, http://dx.doi.org/10.1007/s12116-007-9014-4.

Hopfenbeck, T. et al. (2013), "Balancing trust and accountability? The assessment for learning programme in Norway: A Governing Complex Education Systems case study", *OECD Education Working Papers*, No. 97, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k3txnpqlsnn-en.

Lepi, K. (2013), "How one school uses social media to empower parents and students", Edudemic, webpage, www.edudemic.com/social-media-parents-students/ (Accessed 20 March 2014).

Lipton, D. and J. Sachs (1990), "Creating a market economy in Eastern Europe: The case of Poland", *Brookings Papers on Economic Activity,* Vol. 1, http://dx.doi.org/10.2307/2534526.

Loeb, S. and D.N. Plank, (2008), "Learning what works: Continuous improvement in California's education system", *Policy Brief*, Vol. 8/4, Policy Analysis for California Education, Berkeley and Stanford.

Mason, M. (2008), "What is complexity theory and what are its implications for educational change? ", *Educational Philosophy and Theory***,** Vol 40/ 1, pp. 35-49.

Ministerie van de Vlaamse Gemeenschap [Ministry of Flemish Community] (2009), *Evaluatierapport van de Proeftuinprojecten (2005-2009)* [Assessment report for the experimental projects 2005-2009], Ministry of Flemish Community, Brussels.

Muralidharan, K. and V. Sundararaman (2010), "The impact of diagnostic feedback to teachers on student learning: experimental evidence from India", *Economic Journal*, Vol. 120/546, pp.187-203.

O'Day, J.A. (2002), "Complexity, accountability, and school improvement", *Harvard Educational Review*, Vol. 72/ 3, www.hepg.org/her-home/issues/harvard-educational-review-volume-72-issue-3/herarticle/_48.

OECD (2014), "The weight of nations", *Trends Shaping Education Spotlight 2*, OECD Publishing, Paris, www.oecd.org/edu/ceri/SpotlightBodyandSociety.pdf.

Page, S.E. (2006), "Path dependence: An essay", *Quarterly Journal of Political Science*, Vol. 1/1, pp. 87-115, http://dx.doi.org/10.1561/100.00000006.

Popper, K.R. (1957), *The Poverty of Historicism,* Routledge, London.

Sachs, J.D. (2006), *The End of Poverty: Economic Possibilities for Our Time,* Penguin Group, USA.

Schwerdt, G. et al. (2011), "Effects of adult education vouchers on the labor market: Evidence from a randomised field experiment", IZA, Bonn, www.ideas.repec.org/p/iso/educat/0094.html.

Snyder, S. (2013), "The Simple, the complicated, and the complex: Educational reform through the lens of complexity theory", *OECD Education Working Papers*, No. 96, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k3txnpt1lnr-en.

Swedish National Agency for Education [*Skolverket*] (2011), "Municipal responsibility in practice", Swedish National Agency for Education, Stockholm.

Taylor, C. (2009), "Choice competition, and segregation in a United Kingdom education market", *American Journal of Education,* Vol. 115/ 4, pp. 549-568.

Thompson, M. and D. Wiliam (2008), "Tight but loose: A conceptual framework for scaling up school reforms", in C. Wylie (ed.), *Tight but Loose: Scaling Up Teacher Professional Development in Diverse Contexts*, Educational Testing Service (ETS), Princetown, New Jersey.

Van der Klaauw, W. (2002), "Estimating the effect of financial aid offers on college enrollment: A regression–discontinuity approach", *International Economic Review*, Vol 43/4, pp. 1249–1287.

Van Twist, M. et al. (2013), "Coping with very weak primary schools: Towards smart interventions in Dutch education policy", *OECD Education Working Papers*, No. 98, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k3txnpnhld7-en.

Van Zanten, A. (2003), "Middle-class parents and social mix in French urban schools: Reproduction and transformation of class relations in education", *International Studies in Sociology of Education*, Vol. 13/ 2, pp. 107-123.

Waslander, S., C. Pater and M.V. Weide (2010), "Markets in education: An analytical review of empirical research on market mechanisms in education", *OECD Education Working Papers*, No. 52, OECD Publishing, Paris, http://dx.doi.org/10.1787/5km4pskmkr27-en.

Williams, R.J. et al. (2002), "Two degrees of separation in complex food webs", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 99/20, pp. 12913-12916, http://dx.doi.org/10.1073/pnas.192448799.

Chapter 9.

Experimentalism in Dutch education policy

Lex Borghans, Trudie Schils and Inge de Wolf

Maastricht University

Policy experimentation has the potential to be an effective instrument for policy making in a complex environment. This chapter discusses the experience of the Netherlands, which has engaged in active policy experimentation for the last decade, and distils lessons learned.

Starting with the underlying rationale of policy experimentation in education, the chapter examines the scope of experimentation and innovation in the Dutch education system and describes examples of the various forms of experiments carried out as well as dilemmas and lessons related experimentation. The role of education practitioners, ensuring schools' capacity as well as knowledge dissemination are found as critical for successful experimentation.

Introduction

In recent years the Netherlands has accumulated a great deal of experience with experiments in education. Over the last decade, there has been a shift in thinking about innovations in education, a shift in which experimental research has played an important role. The idea behind experimental research is that it results in a better understanding of what works in education. It has the potential to prevent situations in which educational reforms designed on the drawing board fail to have the intended effects on education once they are introduced. By systematically comparing different ways of organising education, we can obtain a clearer picture of what works in practice, not only in the classroom but in the school as a whole. It is thought that, by taking this approach, we can improve education based on knowledge that has actually been tested in practice. A large number of education experiments have been set up in recent years, making use of a quasi-experimental approach to gain an impression of what works in education.

In this chapter, we describe the Netherlands' experience of setting up experiments as part of its policy on education. We will chart the lessons learned and describe the factors that have led to success or failure when conducting experimental research in education in the last decade. In doing so, we will distinguish between the various phases of a project: the start of the experiment, the execution of the study and the dissemination of the results. Particular attention will be paid to the different perspectives of the various parties involved, including teachers, school heads, governors, administrators and researchers. The chapter also examines the implications of experimental research for the way in which policy is made: what lessons can be learned from Dutch experiences of experimental research in education?

What are experiments in schools?

Experiments in schools are studies in which the effect of an approach (intervention) is examined by comparing a group of pupils who underwent the intervention (the experimental or intervention group) with a group of pupils who did not undergo the intervention (the control group). An intervention can take many forms, such as a new teaching strategy or a new package of educational materials, to name but two. For decades now, experimental research has also been used in other disciplines, such as psychology, economics, criminology, sociology and education sciences.

A particular feature of experimental research is the random assignment of pupils to one of two groups. The randomised division of pupils into intervention and control groups is an important feature of experimental research because it offers a guarantee that both groups are similar in composition. This is of particular importance in education research, since many choices in education result in pupils and students being immersed in a different educational experience and following alternative educational routes. So-called 'selection effects' are always present, and this makes groups of pupils difficult to compare (see also Blanchenay and Burns, Chapter 8, this volume). For example, to discover why some people spend longer in the education system than others, it is not enough to compare people who spend more time in the education system with those who spend less time (e.g. people who complete a university degree and those who enter vocational education). Such a comparison will produce a distorted view of the effects of education since the two groups being compared not only differ in the amount of education they receive, but can also be distinguished by pre-existing differences, such as their ability to learn. A simple comparison of the results achieved by both groups would

therefore reflect not only the effects of additional education but also the difference in people's other characteristics. To effectively measure the effect of education on future outcomes, prospective university students would have to be assigned to vocational education and vice versa. Although often not feasible in practice, this principle is a benchmark for experimental education research. Since education is a very important part of people's lives, it is to be expected that people who have made different choices in education will also differ in other ways. Randomisation is therefore more than a method of ensuring that the intervention and control group have the same composition. It is crucial to eliminate differences that almost certainly exist as a result of selection effects.

From cohorts to experiments in schools

Since the 1970s the Netherlands has established a number of education cohorts, designed to study a large group of pupils over a long period, a process that involved testing and administering questionnaires to pupils, parents and teachers (COOL, VOCL, PRIMA)[1] . This enables researchers who had access to this data to carry out analyses of education at their desks. Experimental research, however, requires a completely different way of working. It involves co-operation with schools, and thus with governors, school heads, teachers and other stakeholders. Support for the experiment among all stakeholders is essential. The intervention being made must be practically applicable, and the pupils at the schools involved in the experiment should be monitored over an extended period. Newer initiatives in data collection within education therefore follow all pupils in a given region, while working more closely with the education sector in order to have a data infrastructure for evaluation when experiments are carried out (e.g. the Onderwijs Monitor Limburg[2]). While the use of experiments to help determine the direction of educational development seems to have great potential to bring about actual improvements in education, such an approach cannot be taken for granted. In the sections that follow this chapter explores some of the challenges and lessons learned from the field.

A comparison with aerospace

When considering a new experimental approach and the changes in education research it implies, parallels can be drawn with the field of astronomy.[3] For a long time, astronomers relied solely on telescopes to study the planets and the stars. Just as an education researcher was able to conduct analyses at his desk, the astronomer did not have to step out from behind his telescope. The transition from traditional research to experimental research in education might be compared to the step from telescope-based research to space flight.

The aerospace sector was not built in a day. To successfully launch rockets and space probes and to carry out measurements across astounding distances, scientists began with small-scale test flights and learned from their failures. Even with all the experience we have accumulated, rockets still malfunction. Or success is only partial, as was the case with the Philae lander: it was lowered onto a speeding comet from the Rosetta space probe but it failed to latch onto the surface securely. In such cases, many years of hard work and dedication result in less information than was originally envisaged. Of course, a great deal can be learned from such setbacks, increasing the next mission's chances of success. This will also apply to experiments in education. It is not reasonable to expect that the introduction of an experimental working method will result directly in perfect interventions. This too will be a process of trial and error.

The importance of experiments

Learning by experimenting

Experimentation is not unique to researchers. Teachers, for example, also experiment a lot. A study by Rivkin, Hanushek and Kain (2005) shows a sharp rise in the quality of a teacher's work during the first two years of his or her career. A likely explanation for this is that novice teachers discover through trial and error how to teach in an effective manner. They notice that what they are doing is not working satisfactorily, so they try a different approach for a few days and, depending on whether they think it is more effective, they switch to the new approach, which then becomes the basis for further experimentation. The ideas for a new approach may well come from colleagues who have drawn upon their own experiences. Every day teachers spend using an approach that later turns out to be relatively ineffective will reduce their productivity. As they increasingly discover what works, this will bring them closer to a more effective way of teaching.

Duration, size and cost of experiments

Not all improvements in education can come from teachers' own experiments. There are two important reasons for this. First, education is all about the pupils' development in the longer term. Most teachers see their pupils for only one year, making it difficult for them to relate the effects of their teaching to outcomes further down the line. In addition, a sufficient number of pupils is needed in order to achieve a meaningful comparison. The magnitude of the effect times the square root of the number of pupils determines the accuracy with which a comparison can be made. This means that a teacher is well able to observe large differences between approaches on the basis of a class of, for example, 30 pupils. For smaller effects – which, incidentally, can still be very substantial – it may be necessary to compare thousands of pupils with one another. A single teacher cannot achieve such a level of comparison, and a more systematic approach is needed.

This makes experimental research in education relatively expensive. Important effects are mostly longer-term effects and the conditions in education are difficult to control, so that the desired effect is often small compared with the variety of influences to which pupils are exposed. Richard R. Nelson and Sidney Winter cite the cost of research as an explanation for the development of science through time (Nelson & Winter, 2009). It is self-evident that researchers will start by analysing correlations for which only a small sample is required. Large effects with few environmental influences are, as it were, the low-hanging fruit that is picked first. That may explain why experimental research first emerged in agriculture, medicine and psychology. Furthermore, in the early stages of research in those areas you can see a particular focus on brief interventions with effects in the short term. In medicine, for example, drugs are tested experimentally, while much nutritional advice is still based on traditional longitudinal research. The benefits to be gained by research also play a role. It is interesting to note that agriculture was making use of experimental research as far back as the early 1900s (Morrison, 1936). Farmers have a strong commercial interest in good farming techniques. In all likelihood, the banks – as party that stood to gain from a farmer's success – also played an important role in encouraging experimental research. Education research is not only relatively difficult due to the long-term nature of expected effects and the strong influence of external factors, but may also be less stimulated due to a lack of commercial pressure.

However, the importance of a good education is increasing at the same time the resources for giving young people more education over a longer period are continuing to

decline, increasing the importance of making the years they spend in education as effective as possible. In addition, the cost of education research is continuing to decrease, most notably due to the emergence of information and communication technologies (ICTs). Many data relating to pupils, in particular test data, have already been entered into computer systems, and the introduction of a single registration number for people in education in the Netherlands (*onderwijsnummer*) is making it increasingly easy to link data to arrive at a bigger picture. On the one hand, it is therefore becoming more valuable to know what does and does not work in education and on the other hand it is becoming easier to monitor pupils over longer periods of time in education. Because the tracking of pupils is especially valuable if there are proper intervention and control groups to address crucial questions in education, this will only increase the value of experimentation.

Scope for experimentation and innovation

Freedom of education is a fundamental principle of the Dutch education system, anchored in the nation's constitution at the beginning of the 20th century after a hard-fought political struggle. Freedom of education means that groups and movements within society are at liberty to establish their own schools and, as long as they meet certain conditions, that these schools are eligible for the same funding as public schools. At the time when this principle was adopted, most of the movements that established their own schools did so on religious grounds. Today the religious identity of a school is far less important to most of the Dutch population, but "special-status schools" are still largely free to design and organise their education as they see fit. In the meantime, the government has been increasing the autonomy given to public schools, creating an overall setting in which schools and school boards have a high degree of educational autonomy while the government generally exercises restraint when it comes to imposing regulations.

Nevertheless, there is often tension between this freedom of education and the government's desire to manage education and encourage improvement. This is perhaps best illustrated by a number of educational reforms implemented in the 1990s, which – in the perception of many – had an adverse effect on educational standards. These were the introduction of a stronger emphasis on independent study in senior years of secondary education (*studiehuis*) and the foundation cycle *(basisvorming*) in the early years of secondary education, and the merging of the basic tracks in general education and vocational education to form a single preparatory vocational secondary education track (VMBO, *voorbereidend middelbaar beroepsonderwijs*; preparatory middle-level applied education). Other examples of government influence on education from that period include reduction of class sizes in primary schools, encouraging ICTs in primary education, and mergers and upscaling in vocational education. The public response to a number of these reforms was rather negative, giving rise to a general mood that the standard of education in the Netherlands was in rapid decline.

A parliamentary committee led by Jeroen Dijsselbloem (Dutch politician) was set up to look into these developments and to identify any lessons that might be drawn from them. One of the committee's conclusions was that, while it could not be said with any certainty that these reforms had led to a drop in the standard of education, it was safe to conclude that the government's control over the situation was not firm enough to prevent a loss of quality. With a view to organising educational improvement more effectively in future to achieve genuinely positive effects, the committee highlighted the importance of gathering good data so that education could be monitored properly: existing evidence should be examined more closely before proceeding to implementation and experiments

would be valuable instruments in this regard. The committee also concluded that the government had been interfering too much with educational reforms and that greater autonomy should be returned to the schools themselves.

In theory, the autonomy of schools and school boards in the Dutch setting creates optimum conditions for variety and scope for experimentation. Yet at the same time, the evidence suggests that these autonomous schools are not always inclined to learn from one other. They sometimes cooperate within the same administrative or collaborative setting, yet at the same time it is clear that dissemination of knowledge about what "works in education" is not an automatic process. The widespread use of ready-made teaching methods and reliance on consultancy firms and organisations is another factor which does not contribute to mutual knowledge-sharing between schools. As regards experimental education research, this is a missed opportunity, since it means that investment in an experiment at a particular school is unlikely to benefit other schools.

Examples of different forms of experiments in the Netherlands

Experimental education research has grown dramatically in recent decades. Experimental and quasi-experimental research (i.e. research in which coincidences that occur in the real world are used as a substitute for an actual experiment) began to be used increasingly as an alternative to the existing methods. What follows is an overview of some key examples from the Netherlands (for a more detailed overview of the development of experimental educational research in the Netherlands, see Borghans, Schils and de Wolf (2016)).

Experiments within the OnderwijsBewijs programme

In 2009, with support from the Ministry of Finance, the Ministry of Education launched a research programme called *OnderwijsBewijs* (Education Evidence), which enabled schools and knowledge institutions to apply jointly to take part in educational experiments on a number of themes. The programme consisted of two rounds. In the first, 18 grants were awarded for the themes of giftedness, language learning and arithmetic, teacher shortages, continuous learning pathways, early childhood education and child welfare. In the second round, 19 grants were awarded for the themes of behavioural problems and bullying, reducing backlogs, excellence and citizenship. These included both projects initiated by a strong impulse from the teaching profession and projects initiated by the world of research. The experiences from the first round showed that while experiments in education are a wonderful idea in theory, putting them into practice was anything but straightforward (see also de Wolf & Borghans, 2012):

Collaboration with universities to build school capacity was met with some reluctance

The most successful projects were those created when a group of schools experiencing particular educational problems enlisted the help of a university to formulate the research question and to design an experiment to determine the most effective approach to the problem. Projects initiated by an individual school tended to encounter problems with the experimental design and face issues of generalizability and scalability. In some instances, misunderstandings arose about the design aspects of experimental research. For instance, some of those involved turned out not to be aware of what a randomized trial was. Interestingly, though a number of project groups were convinced that randomization simply was not possible in their particular case, randomization

ultimately proved possible in almost all of the projects and achievable by means of a design that was acceptable to those involved. Some experiments started with a pilot project. It proved particularly difficult to involve sufficient numbers of schools in university-initiated projects. Level of participation was less of a problem in projects where the research question came from a group of schools, whether working in combination with a university or not.

Problems with the actual execution of experiments can be largely overcome by conducting a pilot project in the first phase of a study

This involves trying out the intervention in a small number of schools. A pilot of this kind is of great value, not least in providing a so-called process evaluation for the experiment. What are you likely to encounter when carrying out the experiment in practice? What solutions are available? There are always issues which neither the teacher/school head nor the researcher have anticipated but which can have a major impact on the execution and results of the experiment. A pilot can prevent disappointment due to teething troubles and may even provide information about the expected effect size, information which can then be utilized in the design of the experiment. Many of the above-mentioned experiments set up from within the teaching profession are in fact closer to pilot projects than full-blown experiments because they are essentially geared towards "trying something out".

The execution of experiments needs to be monitored

Last but not least, it is important to monitor the execution of the experiments. The assumption that an intervention has been made in accordance with the researchers' stipulations often proves to be erroneous. Co-ordination with and the co-operation of schools, teachers and pupils is crucial to implementation. Selective drop-out from the control group is often the biggest source of concern, often caused by a lack of understanding about the intervention.

For an experiment to provide a clear picture of whether an intervention works, it is essential that a sufficient number of pupils/schools participate. In light of this, it is striking that in a large number of projects given the go ahead, the sample size was rather small. Research plans often turned out not to be based on an analysis of statistical power that indicates how large the group of participants should be. Another challenging aspect of experimental design is the protocol used for the control group. Some projects excluded pupils in the control group from every aspect of intervention as much as possible. Amid the complexities of real-world education, comparison with current practice is often easier and more interesting to explore. For example, this enables us to test the impact of a new method or additional teacher-pupil interaction. Comparing this with the traditional method shows the additional effect of the new method. However, this does make it more difficult to prove additional effects and the implications for statistical power also need to be thought through. Since the expected size of the effect is smaller, it means that the research often has to be carried out on a larger group of pupils, classes or schools. Experiments in which the control group continues to use the traditional method are met with fewer objections. In such a design, the members of the control group are not denied anything; their exposure to the new method is only postponed until such times as its effectiveness has been tested.

Experiments and quasi-experiments by the Netherlands Bureau for Economic Policy Analysis (CPB) [4]

In recent years, the CPB has occasionally been involved in evaluations in education. Where possible it has attempted to take an experimental or quasi-experimental approach. These studies also produced interesting experiences with the experimental or quasi-experimental method. It turned out to be very difficult to get schools to carry out experiments. For this reason, the CPB used quasi-experiments which sometimes involved asking the teaching profession to follow procedures that increased evaluability. This tendency to look for quasi-experimental opportunities rather than implementing complete experiments created tension between the questions that one would prefer to have answered and the questions it was possible to answer with the data available.

Sometimes it turned out to be very difficult to emulate a good control and intervention group with the available data. This can be illustrated by a study of the effectiveness of additional supervision and support for underperforming schools in Amsterdam (van Elk & Kok, 2014). Since the municipality applied this intervention to all weak schools in Amsterdam, it was not possible to find a control group within the municipality. The study therefore turned to other municipalities. However, this too proved problematic since it required making a comparable selection of schools in other municipalities, based on the assumption that these municipalities were not pursuing other policies that might have an effect on school performance. The more such assumptions have to be made, the more the quasi-experimental method becomes less rigorous, including the distorted results due to selection issues.

A special case is the study of community schools (*wijkscholen*) in Rotterdam, the effectiveness of which was also evaluated by the CPB. The community school is an initiative whereby pupils who are in danger of falling through the cracks in the system can receive an education to improve their job opportunities or guide them towards another educational programme. Because the initiators of this approach are very much against the random allocation of places at the community school – their philosophy is that every pupil is entitled to use this facility – the CPB decided to make use of the fact that there are only limited places available and that in some cases community schools have to turn pupils away simply because they are full. The control group therefore consisted of pupils who were referred to the community school yet were unable to attend because there were no places available at the time. The question of why this form of selection was seen as less objectionable than the randomized system that was rejected in the first place will remain unanswered here.

Since the ministry was keen to gain an insight into the effects of the community school as quickly as possible, the CPB produced an interim report (Van Elk, 2011). At the time of the interim report, a significant proportion of the pupils were still enrolled at the community school. That made a comparison with the control group difficult. After all, many of the pupils in the control group were no longer in education. When pupils still attending the community school were included in the analysis of how many subjects were in employment or training, the results were bound to show a favourable effect for the intervention group. And if these pupils were excluded from the analysis, it would be difficult to identify a relevant comparison group, as it is not known which pupils in the control group would have still been attending the community school if there had been a place for them. The CPB conducted numerous robustness analyses and despite these problems came to the conclusion that the community school had a positive effect on the careers of the young people who attended it.

Sometime later, the final evaluation followed. This was easier to perform because by that time almost all of the pupils had left the community school (Van Elk, van der Steeg and Webbink, 2013). Remarkably, this evaluation showed no positive correlation between attending the community school and transition to education and employment. While this reversal in findings was no doubt a painful confrontation for those involved in the community schools project, it does provide strong evidence for the usefulness of the experimental approach: a thorough analysis that appeared to come close to replicating an experimental study, nevertheless produced very different results.

In addition to the effects on the transition to education and employment, the final evaluation also looked at the impact of the community school on crime. The study showed that those who attended the community school were in fact more likely to get into trouble with the police. A breakdown of the results into pupils who had been in trouble with the police before attending the community school and those who had no police record showed that the increase in criminal behaviour only applied to pupils who had previously been in trouble with the police. For the other groups the career effects of attending the community school were shown to be beneficial, however, although the observed effects were not significant. That could mean that for some pupils the community school leads to more criminal behaviour, while other pupils experience beneficial effects with no negative effects in terms of criminal behaviour. If the study had been more extensive, or if it had been based on a random assignment of pupils, for example, these favourable outcomes may well have been significant. This shows that seemingly minor details in the design of a study may have greater effects on the reported findings.

Under the heading *Zicht op Effectiviteit* (With a View to Effectiveness), the Ministry of Education commissioned the CPB and Ecorys to come up with experimental or quasi-experimental designs for the evaluation of policies (Van Elk et al., 2011; Briene and Vlasakker, 2011). In a first round, designs were made for all of the ministry's policy areas. In a number of cases it proved impossible to design an experiment that satisfied the so-called gold standard of experimental research. A number of other designs were actually carried out, but no follow-up took place aimed at finding solutions for those policy evaluations that did not seem to fit the existing mould.

Experiments with performance-related pay

In 2011, a new government was formed: a coalition between the liberal VVD and the Christian democrat CDA. One policy measure in their coalition agreement was that resources should be made available to introduce performance-related pay for teachers. Performance-related pay was one of the promising educational improvements recommended in the reports of the CPB. Partly because the Ministry of Education was keen to ensure a support base within the teaching profession and partly because it was not clear what form of performance-related pay would be most effective, the government decided to initiate this process by carrying out a number of experiments. Schools were invited to submit proposals for performance-related pay which, if they resulted in an adequate impact assessment, would be subsidised by the Ministry of Education. Since groups of schools or school boards were being given the freedom to come up with their own interpretation of a performance-related pay programme, it serves as a prime example of how experimentation and a relatively high degree of school autonomy could go hand in hand. By testing the various interventions in different groups of schools, it would be possible to see what worked and what did not. In a traditional intervention only a single implementation of the planned adjustment can be tried out. A conceptual problem

associated with this diverse approach was that the intended experiments would only show the effect of a proposed form of performance-related pay within the schools that opted for that particular form. In other words, if an approach were found to be effective, it would not necessarily mean that the same effects would occur at other schools which adopted the approach. Strictly speaking, this could only be established by means of a randomised follow-up experiment.

Two problems arose in the run-up to these experiments with performance-related pay. Firstly, many of the performance-related pay proposals developed by the schools bore little or no relation to what was known about this subject from the scientific literature. Secondly, in a number of cases the contact between the researchers and the schools involved was far from ideal. The approach that some researchers took to the design of an experiment often turned out to be far removed from the realities of school life. A gulf existed between the researchers' ideas of how an experiment should be conducted and the experimental possibilities that exist within a school setting.

In fact, this represented a collision between two contrasting visions of what constitutes science. On the one hand there was a vision of science as a particular prescription that must be followed in order to produce good research: analyses that follow this prescription are scientific, those that deviate from it are not (see also Cordingley, Chapter 7, this volume). This was the view held by a number of researchers and also by the Ministry of Education. For example, the ministry wrote that use should preferably be made of the 'gold standard'[5] with respect to experimental research, but that compromises should be made where necessary (Van Elk et al., 2011; Briene and Vlasakker, 2011).

Deviating from the standard prescription is therefore regarded as less scientific. However, it is also possible to see science as the attempt to establish a systematic way of trying to establish the existence of certain effects as effectively as possible, given the actual conditions. If we return to the comparison with aerospace research, it is far simpler, for example, to carry out soil analysis on Earth than to do so on a distant planet by sending a space probe. If the analyses on Earth are more accurate than the measurements on the distant planet, this does not mean the latter are not scientific. The scientific challenge lies in developing methods whereby the problems encountered while gathering data on a distant planet are alleviated as much as possible.

This is similar to the challenge educational researchers face with regard to experimental education research. They are familiar with the ideal of a randomized experiment and this forms an attractive prospect for education research. The scientific challenge is to set up experiments in a school context which benefit as much as possible from the power of the experimental approach while coming up with solutions to any problems that arise along the way.

In the end the coalition government was short-lived and the experiments with performance-related pay did not materialize. This was partly because performance-related pay was a highly sensitive issue for the trade unions, one which provoked fierce union opposition. They were convinced that performance-related pay was not feasible and would diminish rather than enhance the motivation of teachers. It might be argued that the doubts surrounding the effectiveness of this instrument made it an ideal candidate for experimentation, but given the alternative logic that governs political processes, it would probably have been better not to initiate the experiments in the first place (see also Burns and Blanchenay, Chapter 10 of this volume, for more discussion on this point).

A subsequent step in the promotion of education on the basis of proven effectiveness was taken in the form of legislation governing anti-bullying policies in schools. In recent years, there has been extensive coverage of the negative effects of bullying in schools, resulting in a political and social consensus that schools should take action to combat it. Legislation was drafted requiring schools to have an anti-bullying policy based on a programme whose effectiveness has been proven by research. A committee was set up to assess whether anti-bullying policies met the requirements. Due to contradictions with the freedom of education principle, proposals for a mandatory effectiveness test were withdrawn. Perhaps solutions to the issue of bullying are too closely bound up with the identity of schools in the Netherlands, most of which are religiously oriented, to permit an approach with such a mandatory component. Mandatory effectiveness assessment may yet become a quality requirement with regard to other themes, for instance as a way of preventing the sale of all kinds of teaching methods to schools without their effectiveness having been clearly established. The experience in relation to this anti-bullying legislation may prove useful in this regard, as coming up with a sound research assessment procedure is no trifling matter. Firstly, such a system should provide ample scope for the effectiveness of untested methods to be evaluated experimentally. Secondly, the anti-bullying issue has generated a good deal of debate about the evaluation criteria to be used.

Other experiments and other challenges

Experimental education research faces another problem: certain data is required in order to carry out experimental or quasi-experimental analysis, yet it is simply not possible to generate or tap into the appropriate data for every interesting question that arises from an educational perspective. This is nicely illustrated by the doctoral research conducted by Ferry Haan[6]. As a journalist and teacher, Ferry Haan has a strong commitment to education and is currently working on a PhD thesis under the supervision of Professor Hessel Oosterbeek at the Univerity of Amsterdam. His study is an attempt to answer questions that he considers important in education and to analyse them in a rigorous manner. This has turned out to be perfectly possible for some questions, while others remain unanswered.

Research into the effectiveness of "Steve Jobs schools" (schools with a strong focus on modern ICT) and IMC weekend schools[7] turned out to be impossible because the problem of selection bias turned out to be insurmountable. No schools of this kind offered quasi-experimental opportunities in the shape of a surplus of applicants, an admission policy based on a lottery or something of that kind. However, such an opportunity was presented by the summer schools pilot project, which provides extra tuition during the summer months to enable pupils to obtain a pass in subjects they failed first time round to avoid having to repeat a year at school. The aim was to make use of the application surplus for summer schools for the purposes of evaluation. Schools volunteered in dribs and drabs but with the summer holidays fast approaching the researchers started with the first schools that had agreed to participate. In the end there were enough participating schools to justify a study, but because places were not allocated randomly, the schools that volunteered early took part while those that volunteered later did not. This led to a form of selection, which meant that the research was no longer possible. It is interesting to note that while many politicians were quick to praise the summer schools as a success, in fact their performance has not yet been the subject of a proper evaluation. The fact that many pupils who attended the summer schools progressed to the next year was seen as a

success but to date we have no way of knowing how many pupils would have achieved the same result without attending the summer school.

Another plan was to analyse the innovations in the teaching of economics at secondary school level. These innovations involved requiring pupils to read more text as part of their programme. The plan was to examine the effects of this change on the differences in performance between boys and girls. However, this turned out not to be the only innovation taking place. As the pilot projects were getting under way, a number of subdivisions within school subjects were abolished, making it impossible to analyse the effects of the content-related innovation in its own right. The one subject which did provide scope for comparison was mathematics. A study was carried out which showed that changes in the teaching of Mathematics B at secondary school (HAVO, *hoger algemeen voortgezet onderwijs*; higher general continued education) resulted in more girls opting for Mathematics A.

In addition to the above-mentioned evaluation with regard to mathematics, it has also proved possible to evaluate a number of other projects. One of these is a regression-discontinuity analysis of excellence programmes at three schools. At one school, the study has been completed and the results are both large and significant. Another analysis is focusing on data from two schools which randomly tested half of their first-year pupils on intelligence, eagerness to learn and their responses to the school questionnaire. The research is focused on whether this will improve the level they achieve beyond first year.

Lessons, dilemmas and opportunities

Interventions and comparability

It is often thought that experimental research cannot make a fair comparison unless the pupils in the control group and the pupils in the intervention group meet exactly the same conditions. This idea is often used to argue that experiments cannot be conducted in an educational setting. After all, pupils are continuously open to all kinds of influences outside the school environment. However, the good thing about randomised experiments is precisely that they ensure that outside influences do not affect the validity of the experiment. Since pupils are assigned to one of the two groups on a random basis, pupils affected by other influences in addition to the intended intervention will be divided approximately equally across intervention and control group. This means that distortion in the comparison between the two groups is avoided. However, the greater the variety of influences that pupils undergo, the smaller the effect of the intervention (measured in terms of relative influence), and the larger the group of pupils required to measure this effect with the same precision. Since the precision depends on the square root of the number of pupils, four times as many pupils are required if there are twice as many differences in environmental influences. While the experimental method is particularly effective in detecting the impact of a specific intervention even if there are many other influences at work, researchers often try to limit these influences as much as possible, so that they can keep the sample size small.

Ethical concerns

Ethical concerns are frequently cited as reasons not to carry out experimental education research. The argument is that it is unethical to provide some pupils with an intervention while withholding it from others. Especially when it is assumed that an intervention will be highly effective, withholding it from a control group is used as an

argument for not using the experimental method to test the effectiveness of the intervention. Of course, it is doubtful whether this conclusion can be reached so simply. If the effect of the intervention is indeed beyond any doubt, it is worth asking why an experiment is needed at all; in most cases it is uncertain whether or not an approach will have the desired effect. It is also not clear why exposing pupils to an untested intervention would be more ethical than testing the effectiveness of an intervention. We have the impression that the ethical argument is often used as a readily available argument against experimental research, when in fact there are underlying objections on other grounds. In projects where all parties are more involved in the design of an experiment, the ethical argument is less likely to surface.

With regard to this particular aspect, experimental education research can learn a great deal from the medical sciences, where researchers are required to identify any ethical concerns beforehand. When such concerns are found, solutions are often sought and submitted to an ethics committee. In experimental education research, this practice is far less common. However, we recommend that it should be adopted, especially when it comes to issues of privacy and enriching research data.

Interestingly, there is sometimes tension between the value of an experiment for a participating individual (pupil, teacher) and its value to society (education, schools). Experimental research can yield a great deal for education as a whole, but for the pupils and teachers who participate, the research is sometimes of less benefit. The intervention is usually withheld from the control group, for instance, which means that the direct value to members of that group can sometimes be disappointing. However, there are ways to address these concerns, such as offering a postponed intervention once the results of the experiment are known (if the intervention proves effective).

Duration, size and cost of experiments

Experimental education research is not always cheap. The cost of small-scale experiments, for example, within a class, is often reasonable. But when the research involves multiple classes and multiple schools, the costs quickly rise. The design and implementation of interventions in educational practice and the measurement of the effects on pupils is especially time-consuming. The drain on financial resources can rise dramatically as a result, especially when the experiments become larger or more complex. It is worth pointing out that in their design and when measuring results, studies do not always make use of existing resources such as measurement data already compiled by the school and pupils' registration data. Yet the use of such data is relatively inexpensive and leads to a huge increase in the usefulness and feasibility of experimental education research. For example, it becomes relatively easy to study the added effects of an intervention, such as effects on pupil performance, school careers and other long-term effects.

The magnitude of an effect is seen by statisticians as the correlation between the effect of the intervention studied and the extent of other influences on the development of the pupil. As a study covers a longer period of time, other influences become larger. Long-term research therefore requires a larger number of participants than short-term research.

Bringing together expertise in experimental research

These experiences of experiments in education bring three types of expertise to the fore. All three are required to set up and carry out an experiment effectively:

1. Knowledge of the education sector: this is knowledge of how educational practice operates and is organised, in the classroom and the school as a whole.
2. Scientific/content-related knowledge: this is knowledge of the theoretical and empirical literature in a specific area.
3. Statistical knowledge: this is knowledge of the design of an experiment, how randomization works and how problems affecting randomisation can be overcome.

Some of the problems setting up experiments in education arise because a single actor rarely possesses these three types of expertise, and there are often differences between how a school and how a researcher approach an experiment. Ensuring that the three different types of relevant expertise come together in a research team can result in an experiment that is workable in the educational setting, the effects of which are attributable to the intervention and the results of which are of both scientific and practical value. Progress has been achieved in this area in recent years. For instance, an increasing number of consortia are being formed between researchers and professionals from the educational setting when it comes to the design and implementation of an experiment. However, at present such consortia are only temporary in nature, a response to the requirements of a given subsidy programme. More sustainable solutions to the experimentation problems outlined above can be provided by long-term partnerships between schools and researchers, giving rise to an ongoing dialogue about the problems that schools experience and what scientists can provide in terms of literature, whereby those involved get to know and understand each other's world (see also Cordingley, Chapter 7, this volume).

Schools participating in experiments

There is a difference between experiments that are initiated on the basis of policy or research and experiments that are the result of an impetus from within the teaching profession. Policy-makers and politicians are often driven by the need to answer questions about whether a particular policy measure is working or not. The problem being studied or the intervention being implemented is not necessarily shared by all schools. How would you then set about determining the participation of schools in such an experiment? There are at least two possible ways to approach this. One way is to determine the participation of schools in research randomly, for example by means of a lottery procedure organised by the Ministry of Education (aside from the matter of which participating schools then become control or intervention schools).

Another approach would be to open up the research to all relevant schools and invite applications to participate. The latter approach best reflects a situation in which schools have the freedom to organise their own education. In that case any effect found will be relevant to those schools alone, as schools that voluntarily participate in the research do not constitute a representative sample of the total population of schools. It could be that some schools are eager to participate in experimental research because they are struggling with the specific problem being addressed in the study. Additionally, it may be that schools object to participating in a study in which the division into intervention and

control groups is determined by random allocation, with no assurance of being part of the intervention group. School heads who decide to participate in a study may differ in the extent to which they are prepared to accept such uncertainty and this may affect participation in experimental research. This prompts an expectation that the findings would say something about how the tested measure/policy would pan out if it were to be applied to *all* schools. However, a disadvantage is that the tested measure/policy need not be useful to all types of schools and thus it would not be useful to involve all schools in the study.

Relationship to data collection

The development of experimental research cannot be viewed separately from the availability of data. Since education is crucial to personal development, the important outcomes of educational interventions only really become visible in the long term. The Netherlands has a long tradition of data collection in the context of education cohorts: since the 1970s, regular studies have taken place which monitor a large group of pupils as they pass through primary and secondary education. For a long time the primary and secondary school cohorts were conducted separately but with the introduction of the COOL cohort study in 2007, an attempt is being made to monitor pupils through both primary and secondary education.

Many educational experiments make no use of such longitudinal data. This means that, in many cases, only the short-term outcomes of specific interventions are examined. Establishing links with the ongoing cohort studies would be difficult, because this implies that the schools participating in the cohort and the schools participating in the experiment (as intervention or control school) would have to be the same.

Statistics Netherlands is increasingly using administrative details as part of its data collection. The organisation has adopted this approach so that all sources of data on individuals can in principle be related to one another by means of the citizen's service number. This will gradually result in a large longitudinal file that covers the entire population. This will ultimately do away with the problem of the group of schools where an experiment is carried out having to be identical to the cohort group. Additionally, this administrative approach enables pupils to be monitored for a very long time.

At present, the administrative data collection amassed by Statistics Netherlands mainly contains details of people's employment situation. In recent years, however, more and more data on education has become available. Some of the above-mentioned cohort studies have been linked to Statistics Netherlands' administrative data, which provides data on the further life experience of pupils who were originally only monitored during their years at secondary school. For the further development of experimental research it would be of great value if, for the schools that form part of the study, data about the further career of pupils were to be made available through this route.

Dissemination of results

A final important lesson to be learned from the range of educational experiments in the Netherlands is that improvements need to be made regarding the dissemination of results. The large degree of autonomy enjoyed by Dutch schools and the lack of knowledge exchange between them means that relatively few schools benefit from the results of experiments. In this regard, too, bridging the gap between teaching practice and scientific knowledge is no mean feat. Given the cost of experiments and the valuable insights they produce, this is a great pity. There are a number of initiatives geared towards

sharing knowledge about 'what works' with professionals from educational practice. For instance, the Ministry-funded TIER-institute, set up in 2010 with the aim to generate evidence-based knowledge about education, has developed and launched a Best Evidence in Education website[8]; there is a website that features best practices and results of experiments conducted as part of the *OnderwijsBewijs* programme[9]; and Marzano has written several books on interventions in education which have been tested and found to be effective (e.g. Marzano, 2003). These are all interesting initiatives, but not widely used within the teaching profession. In this respect, there are clear differences between the situation in the Netherlands and that in the United States, for example, where much greater emphasis is placed on the dissemination of the results of experimental studies (e.g. through What Works Clearinghouse[10], BEE[11], incentives for stimulating effective methods). Here in the Netherlands it is often the government or the academic world that takes the initiative and encourages the dissemination of measures or good examples.

Conclusion

Experimental research pays

Over the last decade, the Netherlands has amassed a great deal of experience of experimentation in education research. Especially in combination with the development of good data collection where pupils' long-term development can be monitored over the longer term, this experience has great potential as regards achieving systematic improvements in education.

The randomised experiment is elegant and simple, yet carrying it out in practice is far from straightforward. We have a long way to go before the question of how best to design experimental research that is effective in educational practice can be fully answered, and it will require much more in the way of scientific creativity to develop intelligent approaches for this purpose. Nor is it the case that well-designed experimental research always leads to irrefutable answers. Details in the design of a study may influence the outcome, as with any type of research. Although policy-makers need clear answers, the power of science must continue to reside in the fact that all findings remain open to discussion. Even conclusions and interpretations that are almost universally accepted at a given point of time can be seen in a new light as a result of new research or new approaches.

Challenges for the future

To ensure the success of experimental education research, constructive co-operation with the teaching profession is crucial. At present, this is a major bottleneck in the development of experimental education research. In an ideal world, schools and school boards would test any changes they plan to make using an experimental approach, before proceeding to implementation. Given that such changes in education are often projects that go far beyond standard research budgets, effective co-ordination between researchers and the plans within the educational setting is of the essence. At present, this aspect is still fraught with difficulty and there is often a lack of co-operation and dialogue between research, policy and the profession. Organizing such co-operation is also a complex matter. But if our aim is to use experiments to answer bigger questions and look at long-term effects, co-operation must amount to more than collaboration on a one-off experiment in one or several school classes. Such co-operation is not easy to organize; it requires investment, perseverance, scope for experimentation and good mutual relationships. As yet, such partnerships and relationships between teachers and schools,

scientists and policy-makers are few and far between. However, there are a number of fledgling partnerships in the Netherlands, and this gives us cause for hope. Examples from the United States show that this form of co-operation is not only possible, but also leads to important new insights for education.[12]

A second challenge concerns the selection of themes and interventions to be examined. Experimental research provides insight into the effects of interventions, but does not answer the question of what research needs to be carried out for the further improvement of education. Nevertheless, important choices need to be made in this respect. The matter of how to make such choices and how to facilitate experimental inquiry into important questions that need to be addressed represent a major challenge for the development of experimental education research. If we expect our schools to base their policies on what is known to work and if we assess the plans of political parties on what is known about their effectiveness, it is important that the knowledge they need in order to take such decisions continues to be generated. At present, choices with regard to experimentation are too often dependent on available data, policy themes or ad hoc questions from schools. As a result, some experiments are not designed as well as they should be, some important experiments are never carried out at all and other experiments focus on interventions that have no prospect of ever being implemented by teachers or schools. When drawing up a solid research agenda for experimentation, it is advisable to seek out or bring about co-operation between the teaching profession, the academic world and the policy-makers. It is also important to base choices on knowledge about effects in education, both national and international.

A third and final challenge is to organize the learning ability within teaching. An important step in this direction would be to improve the dissemination of the results of experimental education research. This could reduce reluctance among schools to take action and make them less dependent on incidental choices from the world of educational advisors. An important precondition is that teachers and schools should be given access to scientific publications and other sources of research data. Investments can also be made in review studies and websites that increase the accessibility of results for schools, following the example set by countries such as the United States.

Yet greater dissemination alone is not enough. It would also be highly beneficial if schools themselves were to experiment more and go in search of research partners for this purpose. This process could be more readily facilitated, not only in time and money but also by means of legislative scope. Policy-makers can also do much more to ensure that schools mainly use effective programmes, for example by expecting this of them or by providing incentives to do so. Parents and pupils/students can also make demands in this regard. In addition, education advisors can perhaps play a more active role when it comes to stimulating the learning ability of schools. In the long term, it would be wonderful if schools and the academic world joined forces to take responsibility for the tradition of experimental education research, doing away with the need for government involvement altogether.

Notes

1. Cohort Onderzoek Onderwijs Loopbanen (COOL), Voortgezet Onderwijs Cohort Leerlingen (VOCL), Cohort Onderzoek Primair Onderwijs (PRIMA).
2. The Limburg Education Monitor is part of a cooperative project between Maastricht University and schools, school boards and government bodies in Limburg – a Province in the South of the Netherlands – that aims to further improve education in the region based on systematic data collection, dialogue and (experimental) research.
3. This comparison is also made in Borghans, 'Kunnen we meer leren over leren? [Can we learn more about learning?]', a lecture given before the Education and Labour Committee of the Social and Economic Council of the Netherlands (SER), The Hague, October 2007.
4. CPB, Centraal Planbureau, www.cpb.nl/en
5. Randomised controlled trials (RCT) experiments are often seen as the "gold standard" (a score of 5 on the Maryland Scientific Methods Scale (Sherman et al., 1997). It is used across a range of disciplines, most prominently in the medical sciences.
6. Economics teacher secondary education at Jac. P. Thijsse College, researcher University of Amsterdam and associated member of the Dutch Education Council.
7. Nine IMC (International Market makers Combination) weekend schools provide supplementary education for children aged 10-14 from disadvantaged neighbourhoods, see www.imcweekendschool.nl/home/english/
8. www.bestevidence.org
9. www.onderwijsbewijs.nl
10. The What-Works-Clearinghouse, managed by the US Institute for Education Sciences, identifies scientific studies about the effectiveness of education practices and disseminates summaries and reports (http://ies.ed.gov/ncee/wwc/)
11. www.tierweb.nl/bee/tier-bee.html
12. For example Centre of Educational Policy Analysis (Stanford University) and Educational Innovation Laboratory (Harvard University).

References

Borghans, L., T. Schils and I. de Wolf (2016), *Experimentalism in Dutch Education Policy: Experiences and Lessons Learned*, Working paper Academic Workplace Education, Maastricht/Utrecht, www.academischewerkplaatsonderwijs.nl.

Briene, M.F.M. and S. Vlasakker (2011), *Zicht op effectiviteit van beleid: evaluatieontwerpen beleidsinterventies cultuurdomein* [A view on effectiveness of policy: evaluation designs of policy interventions in the cultural domain], Ecorys, Rotterdam.

Marzano, R. J. (2003), *What Works in Schools, Translating Research into Action*, ASCD, Alexandria.

Morrison, F.B. (1936), *Feeds and feeding: A handbook*, Morrison Publishing Company, Ithaca.

Nelson, R. and S.G. Winter, (2009), *An Evolutionary Theory of Economic Change*, Harvard University Press.

Rivkin, H. and J. Kain (2005), "Teachers, schools, and academic achievement", *Econometrica*, Vol. 73/2, pp.417-458.

Sherman, L.W. et al. (1997), "Preventing crime: What works, what doesn't, what's promising", Report to the U.S. Congress, Department of Justice, Washington, D.C.

Van Elk, R, and S. Kok (2014), "Het effect van de Kwaliteitsaanpak Basisonderwijs Amsterdam op leerlingprestaties: Resultaten van de eerste vier jaar" [The effect of the quality approach in primary education on student performance: Results of the first four years], *CPB Discussion Paper No. 264*, The Hague.

Van Elk, R. (2011), "Evaluatie wijkscholen Rotterdam" [Evaluation community schools Rotterdam], CPB Memorandum, The Hague.

Van Elk, R., F.-B. van der Meer, M. van der Steeg, and D. Webbink (2011), "Zicht op effectiviteit van beleid: Studie naar evaluatieontwerpen voor onderwijs- en wetenschapmaatregelen" [A view on the effectiveness of policy: Study of evaluation designs for education and science policies], CPB Background Document, The Hague.

Van Elk, R., M. van der Steeg, and D.Webbink (2013), "De effecten van de wijkschool in Rotterdam op onderwijsdeelname, werk en criminaliteit" [The effects of the community school in Rotterdam on education participation, work and crime], *CPB Discussion Paper No. 241*, The Hague.

Wolf, I. and L. Borghans (2012), "Ervaringen OnderwijsBewijs" [Experiences of OnderwijsBewijs], in I. Waterreus, and I. Van der Heul, (eds), "Stapsgewijze verbeteringen in het onderwijs en samenwerking tussen onderwijsonderzoek en onderwijspraktijk" [Gradual improvement in education and in the co-operation between education research and practice], *Pedagogische Studien*, Vol. 89, pp. 377-387.

Chapter 10.

Learning to fail, not failing to learn*

Tracey Burns, *Centre for Educational Research and Education, OECD*

Patrick Blanchenay, *University of Toronto*

Education systems must continuously evolve and improve in order to meet the diverse needs of today's learners. This urge for innovation and improvement is at the top of policy agendas across the OECD. However, innovation can be inhibited by governance systems that seek to minimise risk and errors. Although an important element of accountability, they can also inadvertently serve to entrench the status quo. Innovation in education requires careful risk-taking and the accompanying possibility of failure. This chapter discusses two ways in which this can be accomplished: Through experimentation, i.e. the testing of innovative programmes in a limited magnitude and scope; as well as by developing a governance system that can learn from failures as well as successes. This chapter argues that both are useful and necessary elements of a modern, evolving governance system, and provides a brief overview of how each of these two elements might play out in modern education.

* *With apologies to Mark Cannon and Amy Edmondson.*

Introduction

Improving the functioning of public services requires innovation, and the ability to change and evolve with new circumstances and challenges. Innovation in turn requires risk-taking – trying something new, and possibly, failing. In education, there is a push to make our systems more innovative and our teachers more creative. Meeting the diverse needs of today's learners is a requirement of modern education governance, as is preparing our systems to solve the problems of tomorrow (Burns and Weatherby, 2014). And indeed, there are numerous examples of innovative practices in pedagogy and practice that can be found throughout the OECD (OECD, 2015). Yet we must do more: our governance systems must be able to innovate and improve along with the school systems themselves

Yet making this happen is no easy task. Countries must encourage innovation in their education systems at the same time as their accountability systems seek to minimise risk and error (Brown & Osborne, 2013). This is an important and difficult tension: countries are under strong internal (and at times external) pressure to strengthen their accountability systems at the same time as they seek to encourage innovation. Reconciling risk and accountability constitutes a demanding but feasible challenge.

Brown and Osborne (2013, following Renn, 2008) contrast three different approaches to risk management in public services as they pertain to innovation:

1. The *risk minimisation* approach identifies risk as something entirely negative that should be avoided if possible; it neglects the fact that risk is a prerequisite condition for innovation.
2. A *risk analysis* approach recognises risk as unavoidable and tries to minimise its consequences. Brown and Osborne (2013) argue that such an approach is often not fit for purpose in complex environments, with multiple points of views, and might result in a low common denominator, yielding too little innovation. They argue instead for:
3. a third approach called *transparent risk governance*, in which risks are openly acknowledged to all relevant stakeholders, and which "allows the articulation, negotiation and (potential for) resolution of the often contested views about the outcomes of innovations in public services" (p. 198).

This third argument is an important one, providing a model for how public services could manage risk and yet still leave room for innovation. Yet it is not the model that is most frequently found across education systems. Too often education systems remain stuck in a paradigm of risk minimisation, or spend a great deal of time on risk analysis without an accompanying plan for how it might be governed or managed in a positive way.

While understandable, these approaches and the systems that model them are missing not only a piece of the puzzle, but a fundamental precondition to innovative and excellent public service. Traditional approaches which seek to minimise, or simply contain risk not only do not allow for innovation and change, they also ignore a fundamental truth: that the *status quo* can be risky to maintain. No change is also a decision, and one that carries consequences for all stakeholders, first and foremost for students. What is the cost of inaction, or of not adopting a better method/strategy/approach to teaching and learning?

Often the answer is that this cost is simply not known, or not calculated. While this might be politically expedient (and the safest path), it transfers the risk and the costs of inaction or failure to the students and future students.

This chapter will explore two potential ways in which countries and education systems can move towards creating a useable model of transparent risk governance. One way to build risk-taking into the system is through experimentation, which seeks to limit the risks involved in any new programme by rolling it out in a limited manner (see Blanchenay and Burns, Chapter 8). Experimentation offers a systematic approach to understand what can be improved, and, in contrast, a way to highlight the opportunity cost of inaction.

A second way forward is to work on the governance system itself, to develop a system that can learn from failures as well as successes. This chapter will give a brief overview of how each of these two elements might play out in modern education systems.

Experimentation

The notion of risk lies at the heart of experimentation, as it represents an attempt to fill a gap in our knowledge of what works and what does not. In order for educational experimentation to work, education systems must adopt an attitude of constructive scepticism that acknowledges the risk inherent in any reform or experiment and allows them to transparently govern this process. However, making this happen in complicated (and often complex) political contexts is no easy task. Elements that are important to consider include the:

- *Governance of risk:* who is responsible, who decides on the levels of risk appropriate in decentralised systems?
- *Identification of risk:* who or what is at risk and how might this affect different stakeholders?
- *Level of acceptable risk:* this will vary across different stakeholder groups, and also different cultures and contexts.

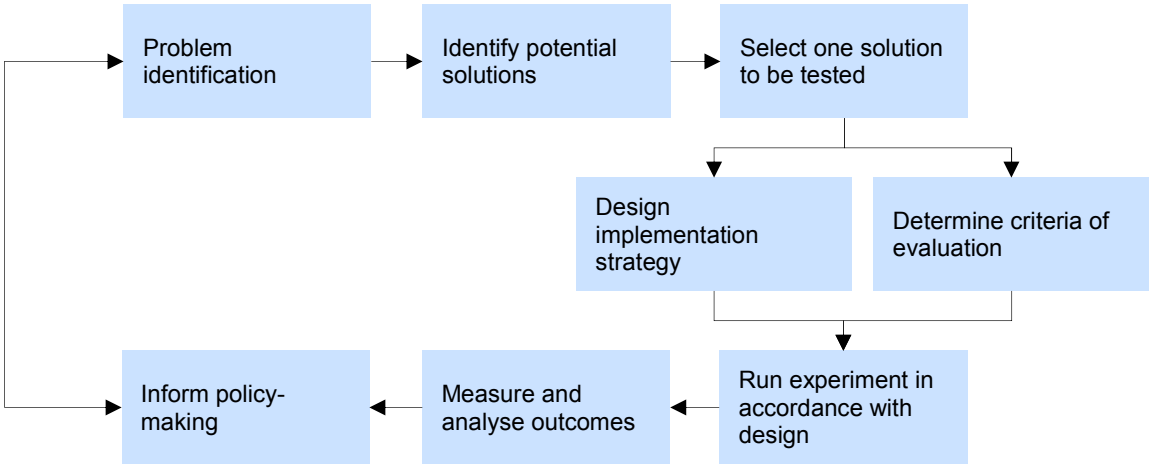
Identifying and acknowledging risks

The first risk of experimentation is an operational one. Once a problem has been identified, each experiment should be thought of as a project in itself from inception to completion (see Figure 10.1 for basic steps). It must involve a design phase in which potential solutions (or treatments) are suggested and sorted and evaluation criteria are decided. Implementation must then be carefully monitored and managed to make sure that what happens at the grass-roots level corresponds to the initial design of the experiment and its evaluation goals. The operational success of an experiment implies that all stakeholders (school staff and parents, local and central authorities, communities, and of course the students themselves) play their parts at all the steps, including the evaluation, which should be performed with respect to criteria decided *a priori*, and if possible performed by independent evaluators.

This *operational risk* impacts all stakeholders involved in the experiment as each of them are required to invest a significant amount of resources for the success of the experiment, whether financial or non-financial (such as time, effort, etc.). Each of these investments is necessary but not sufficient. In this string of decisions and actions, the lack of effective involvement of any given party can affect the success of the whole project.

As such, this operational risk suffers from a co-ordination problem: each actor's effort is only going to matter as long as other actors also put in their share of the effort. The level of this problem will of course depend on the size and extent of the experiment, as well as the number of actors involved (and so, likely to be greater for more highly decentralised systems).

Figure 10.1. Schematic steps of an experimental project



Experimentation also involves a risk related to the actual findings of the experiment, what one might call a *substantive risk* (or content risk): the experiment may identify that the policy being tried out does not have the effect that was initially expected or that it may have unintended consequences which make the policy unsuitable for wider roll-out. For instance, Fryer's (2011) use of financial incentives to directly reward higher test scores in New York and Chicago proved to be ineffective, as they did not result in significant improvement of student performances. These kinds of results, while providing extremely valuable information for policy makers, can be difficult from a political point of view if, for example, the public perception may be that the budget used for that experiment was wasted.

Substantive risk can similarly be political risk if the outcome of certain experiments contradicts previously held beliefs about the effectiveness or suitability of given types of policy (see Box 10.1 for an example). For instance, there is now ample evidence that indicates that smaller class sizes are not associated with higher student achievement (Chingos and Whitehurst, 2011), despite long-standing and deeply held beliefs to the contrary. The risk in this case is not only that the finding might or might not be difficult or unsuitable, but also that there is extra work required to debunk strongly held beliefs (also referred to as "myths"), regardless of whether or not they are supported by clear evidence. This creates a political disincentive to freely reveal risks or failures during an experiment or a policy roll-out. Although understandable, this is a missed opportunity to learn from what went wrong and correct and acknowledge the elements that are most likely to need change.

Political risks can also be considered a separate kind of risk, sometimes called *reputational risk* (Hood, 2002). This added level of complexity is based on the simple truth that politically elected decision-makers have clear disincentives to avoid being

blamed for risk and failure. This contributes to risk-avoidant decisions (e.g. choosing to remain with a minimally disruptive status quo if the results are not too dire rather than seek to improve, which would require some risk and thus potential failure). This is an extremely strong pressure: scholars in social psychology and political science have focused on the "negativity bias", in which individuals have a preference for or bias to remember negative effects and results over positive ones (see, for example, Kanouse and Hanson, 1972; Lau, 1985; Weaver, 1988).

While these results apply to individuals generally, the effect is likely to be more pronounced in those dependent on public perception and approval for their continuing professional well-being. Weaver (1986/8), for example, has argued that American politicians are motivated more by the desire to avoid blame for negative outcomes than by the desire to claim credit for positive ones. This is clearly not just an American phenomenon, and in today's non-stop mediatised climate, seeking to avoid blame is likely to be an essential political strategy. Returning to our discussion on experimentation in education, this pressure creates a situation where experimentation is not only potentially risky for operational purposes, but also for political ones. There is thus an incentive to a) avoid risk, potentially by inaction or remaining with the status quo and b) shift or hide negative interpretations of failure so as to avoid any political or reputational backlash. This last element is a key barrier to creating systems that can learn from failure, an issue that will be addressed in the second half of this chapter.

Paradoxically, such reputational risks are more prevalent if the evaluation of the experiment is done rigorously and by independent parties as there is then less likelihood that the results of the experiments might get massaged, or that the evaluation might be designed in a way to favour certain outcomes. It is precisely when there is a risk of skewing the results of an experiment that proper rigorous and impartial evaluation should be carefully designed, to shield experiments from political interference or misinterpretation.

Box 10.1. An example of political risk

The voucher experiment conducted by Schwerdt et al. (2011) in Switzerland illustrates an example of risk, namely that of political risk. The experiment was initially mandated by the Swiss government to find out whether the financing of adult learning should change, and if so, what instruments should be used. The randomized experiment concluded that the average effect of vouchers was negligibly small; however, it did find that vouchers were mostly taken up by already educated workers who did not benefit from them, while the take-up rate was low among sub-populations who benefit more from them.

The experiment suggested that by targeting such sub-populations, vouchers could be an effective tool to promote adult learning. The Swiss government therefore proposed that targeted vouchers be used for the financing of adult learning. However, the proposal was opposed by the Swiss parliament, who decided to use tax deductions instead. While the experiment was conducted properly and fully, its conclusions seem to have run counter to the prevailing political agenda of elected officials at the time (another example of substantive risk as set out in Table 10.1).

Different risks for different actors

Both from an ethical and a practical point of view, it is crucial to properly understand the types of risks stakeholders are exposed to, just as it is necessary to understand the impacts that being part of an experiment might generate. It is also necessary to

acknowledge that this might affect relevant stakeholders' willingness to participate or lead to greater attrition of participants (see Blanchenay and Burns (Chapter 8, this volume) on the problems of attrition).

A further complication with experimental risks in education lies in the specificity of the public sector. In the private sector, it is often possible to design contracts in a way that can allocate risks appropriately between parties. For instance, it is common for a manufacturer to include penalty clauses in case a supplier does not deliver certain required inputs in time; such clauses are designed to shift some of the risk (that of not being delivered in time) towards the supplier, instead of the risk being borne solely by the manufacturer. The possibility to share the burden of risk enables such risky projects to go ahead.

However, such sharing is not always possible in the public sector. For example, some professions may be protected (e.g. in education teachers might be considered civil servants and as such be sheltered from measures that could endanger their employment or pay). Moreover, side payments and monetary compensations might not always be feasible, for budgetary, legal or ethical reasons. For example, teachers taking part in an experiment may not be permitted to receive monetary or non-monetary compensation as that might violate the equity principle as it excludes those teachers that do not participate in the experiment (through no choice of their own). Importantly, teachers could of course volunteer to take part and thus in some areas receive compensation, but then the selection of the teachers themselves would not be random and may result in a bias.

Similarly, parents may find it unethical that their children could be randomly allocated into a control or a treatment group[1], or that an experiment is only available to certain schools or districts in a region. This can affect participation and support by stakeholders and further endanger the validity of the experiment. Mitigating the risks for all parties involved is thus important from an ethical standpoint as well as a scientific standpoint. Table 10.1 provides examples of how different types of risks might play out for the various education stakeholders.

Table 10.1. Risks of experimentation to various education stakeholders

	Operational risks (experiment does not run properly)	Substantive risks (experiment yields unexpected results)
Elected officials	• Wasted budget • Risk of conflict with other stakeholders	• Policy does not fit political agenda • Policy does not please constituents
School principals	• Wasted use of resources • Unnecessary reorganization of processes • Teacher/parent/student discontent	• Conclusion of research might anger teachers or parents • Might suggest important changes are necessary in school organisation
Teachers	• Experiment requires extra workload • Loss of time that could be otherwise spent delivering the curriculum	• Might prove detrimental to students or to themselves
Unions	• Experiment requires extra workload • Teacher/parent/school board discontent	• Ideological disagreement with conclusions
Parents and students	• Change in learning routines • Loss of time that could be otherwise spent	• Might prove detrimental to students
Researchers	• Budget and time investment lost • Reputational damage regarding ability to run experiments • Involvements of other stakeholders might affect methodology and rigour of the experiment	• Might disappoint funding bodies • Might yield an academically unattractive result

Learning from failure: Building constructive scepticism into the system

In the face of the risks highlighted above, it is crucial for education systems to anticipate those risks, both in terms of establishing a process for governing risk and developing a transparent and reactive way to make decisions about the kinds of risks that are acceptable in any given situation. But there is another element of the process that is equally important: education systems must accept that taking risks (in experimentation and indeed in any kind of innovation) means that there is a possibility of failure. This cannot be avoided, and in fact it would be unwise to minimise this possibility, both in the public discussion surrounding policy choices and in the reaction to a failed initiative. Although hiding failure (or rebranding it as a success) is all too human and indeed politically expedient, it is in fact a missed opportunity. Failure can and should be used as a learning tool, both for scientific purposes (understanding what works and what does not) and for political ones (resources can be wasted if the appropriate lessons are not drawn from failures).

A key lesson of complexity theory is the idea that systems may react differently from what is initially planned at the onset of an experiment. Patterns of behaviour and certain dynamics may emerge unexpectedly as a result of potentially minor changes and generate results opposite to the initial hypothesis behind the experimentation. In this context, it is vital to build room for feedback – all feedback, including the negative – to be taken into account.

This can ultimately help the policy experimentation process in two ways. Firstly, it can help limit the risk inherent in experimentation by enabling stakeholders with relevant experience to clearly identify experiments that are bound to fail. Secondly and perhaps more importantly, reforms are often performed with the assumption that they will succeed (Campbell, 1969). But the goal of experimentation is to verify whether a policy suggestion could work in practice. Such a discovery process requires the possibility to identify failure, i.e. what does not work. "Zero results" are results nonetheless; an important part of learning through experimentation is done through refutability, eliminating techniques and potential policies that are demonstrably ineffective.

Experimentation and policy reform more generally aims at the elimination of mechanisms that can be shown not to work just as much as it aims to reinforce those that do. Therefore, it must be conducted with the tacit acceptance that learning that a certain policy would have zero impact is still useful. Although it is politically harder to justify running experiments that in the end yield zero results, it is important to be able to convey the message that it is better to discover this at the experiment stage, rather than once the policy has been rolled out more widely. This can prove a solid argument for policy makers to defend experimentation publicly.

For experimentation to succeed as a process (including learning from possible failure), it is important that potential experiments are discussed openly, both before their implementation and after their evaluation. For this, it is necessary to create a culture of co-operation between all relevant stakeholders, such that, when failures – inevitably – occur, a dispassionate discussion can take place, with the objective of identifying the reasons for failure to avoid or amend them in the future, rather than laying the blame on specific individuals. While this may seem far-fetched, it is not impossible: such endeavours are used successfully to promote innovation in private companies (see Box 10.2).

Box 10.2. Encouraging constructive failure at Google

Google explicitly promotes failure as a crucial part of its innovation process. According to A. Teller, head of their research lab "Google X": "You must reward people for failing. If not, they won't take risks and make breakthroughs. If you don't reward failure, people will hang on to a doomed idea for fear of the consequences. That wastes time and saps an organisation's spirit […] I am asking them to be responsibly irresponsible. I am asking for each of the project, for each of the group to explore, to take risks, to run experiments, to learn from them and to repeat." (Grossman, 2014).

Risk-taking and experiencing failure is not an objective *per se* at Google but identified as a necessary step towards innovation and improvement. Vinton Cerf, senior executive at Google, explained in a 2011 interview: "The reason that we have success is that we do tolerate failure. […] People are allowed to come back and say 'it didn't work'. The important part is understanding why it didn't work. If we understand that, we can make something better because we discovered how not to do it" (Dice.com, 02 June 2011) This approach to risk-taking and embracing failure can be adapted to all systems seeking to innovate and improve. In policy-making, progress can only be achieved if weak policies can be identified as such and replaced by stronger ones.

Even considering that the public sector is considered generally more risk-averse (Osborne et al., 2015), the ability to create a climate of constructive criticism would enable policy experimentation to become a vital tool of policy-making for complex systems. While in the private sector there might be real incentives for companies to use failures as part of their R&D process, the necessity of critical discussion is all the more relevant in education systems, which in many OECD countries tend to be protected from the type of market forces that eliminate unsuccessful ideas in the private sector.

Some interesting examples of how this might work in public sectors are explored by Osborne and Flemig (2015) in their work on risk and social innovation. They build on the seminal distinction between known and unknown uncertainty (Knight, 1921) to adapt it to the governance of risk context. They distinguish between *uncertainty* (which is, by definition, unknown and can lead to spontaneous and unplanned risks and innovations in a complex system) and *known risks* (which can often be quantified and, if not mitigated, planned for). They argue that it is important to make this distinction in order to best think about how to work with these risks and harness their potential for innovation.

Known risks can drive or enable innovation to find new ways of harnessing or controlling these known risks (Osborne and Flemig [2015]) give the example of new waste management techniques in environmental sustainability, or new mental health medications). However, known risks may also be barriers to innovation if the risk is considered too serious (either substantively or politically). In contrast, uncertainty can either freeze or stimulate innovation through sudden shocks. These observations can be combined with approaches to governance and risk management to suggest the framework presented in Table 10.2.

Table 10.2. Managing risk in social innovation

Type of governance/risk management	Risk	Uncertainty
Hard approach (regulation, rules)	Evolutionary Innovation (Top-Down Management)	Stagnation (Risk Minimisation)
Soft approach (communication, adaptation of organisational culture)	Expansionary Innovation (People-Driven Risk Governance)	Total Innovation ("Thriving on Chaos")

Source: Osborne and Flemig, 2015.

How would this work?

When analysing how organizations can harness failure as a valuable learning tool, Cannon & Edmondson (2005) identify three key processes, in order of difficulty: (1) identifying failure, (2) analysing failure and (3) deliberate experimentation. They argue that both technical and social barriers to these processes must be removed. Technical barriers include relevant data collection and the ability to analyse the data as required in order to identify the underlying causes of failure. Social barriers are more subtle and difficult to address, and include an unwillingness to admit to failure, the temptation to attribute failure to other sources, and organisational structures that punish individuals for reporting failure. These are related to, but not entirely the same, as the political risks previously discussed in the chapter.

To facilitate the *identification of failure* they recommend putting in place solid data collection and blameless failure publicizing. As an example of this, they cite how Minneapolis Children's Hospital set up a "blameless reporting" system, involving Focused Event Studies where both major and minor incidents are reported and analysed, as well as cross-specialties Safety Action Teams that are in charge of pre-empting them.[2] Here it is important to precise that the reporting is in this case internal, and the mechanisms are internal as well. It is far less likely to succeed if the reporting is widely publicised and picked up by the media or other actors.

Cannon and Edmondson (2005) argue that *analysing failure* is a crucial (and generally overlooked) element. Here social barriers, such as denying responsibility for a problem, can hinder the ability to learn from mistakes. In order to create a culture where this learning is possible, they recommend explicit "after event" reviews, making experts available and building capacity among staff to discuss and analyse what went wrong and why. This requires leaders who are willing to "walk the walk" and set the example. It might also require expert help to ensure that individuals are not blamed for mistakes, but rather the situation is used as a learning tool, with open and transparent discussion. Examples of this in education can be found across OECD countries (OECD, 2015), and often centre on teacher's professional collaboration, peer mentoring, and coaching (for concrete examples, see also Jensen et al., 2016).

Finally, to encourage *deliberate experimentation*, they recommend that explicit goals of failure rate be set, and that key personnel be trained to serve as internal consultants to manage these experimentation processes. Designing such deliberate experiments is a highly complicated task, but key to asking the right questions that will allow for identifying successes and failures. This analysis of organisational innovation can also be used in educational systems. Table 10.3 provides a summary of their argument:

Table 10.3. Enabling organisational learning from failure

	Identifying failures	Analysing failures	Experimentation
Technical Barriers	• Complex systems make many small failures ambiguous.	• A lack of skills and techniques to extract lessons from failures.	• Lack of knowledge of experimental design.
Recommendations	• Build information systems to capture and organize data, and ensure availability of systems analysis expertise.	• Structure formal guidelines for effective analysis of failures, and ensure availability of data analysis expertise.	• Train key individuals in experiment design; use them as internal advisors.
Social Barriers	• Threats to self-esteem inhibit recognition of one's own failures, and corporate cultures that 'shoot the messenger' limit reporting of failures.	• Ineffective group process limits effectiveness of failure analysis discussions. • Individuals lack efficacy for handling 'hot' issues.	• Organizations may penalize failed experiments, inhibiting willingness to incur failure for the sake of learning.
Recommendations	• Reinforce psychological safety through blameless reporting systems, training managers in coaching skills, and by publicizing failures as a means of learning.	• Ensure availability of experts in group dialogue and collaborative learning.	• Conduct experiments in key areas, and publicize results, positive and negative, widely within the organisation. • Set target failure rate for experiments in service of innovation and make sure reward systems do not contradict this goal.

Source: Adapted from Table 1 in Cannon and Edmondson (2005).

Conclusion

This chapter explored risk-taking in education. In order to innovate or try something new, the risk of failure is always present. Traditional governance approaches which seek to minimise, or simply contain, risk not only do not allow for innovation and change, they also ignore the fact that the *status quo* can also be risky to maintain. No change is also a decision, and one that carries consequences for all stakeholders, first and foremost for students. What is the cost of inaction, or of failing to adopt a better method/strategy/approach to teaching and learning? Often the answer is that this cost is simply not known, or not calculated.

This chapter looked at two possible approaches to support careful risk-taking in education. One is through experimentation, which seeks to limit the risks involved in any new programme by rolling it out in a limited manner. Experimentation offers a systematic approach to understand what can be improved, and, in contrast, a way to highlight the opportunity cost of inaction.

A second way forward is by developing a governance system that can learn from failures as well as successes. This can be at the individual school level or at the system level (whether local or national). Although difficult and challenging, the chapter argues that modern education systems must be able to build learning from failure into their functioning, both to improve pedagogy and practice and the governing of the system as a whole. Yet there are a number of system-wide elements that tend to encourage risk-

avoidance, for example ethical issues of experimenting with children, political issues of admitting failure, and the prevalence of the "blame-game", which provides a direct incentive to cover up failures.

The governance of risk-taking on a system level requires the involvement of many interdependent stakeholders with different cultures, motivations and time horizons, in a careful exercise of co-ordination towards a common goal of identifying which policies work and which do not. As part of this, being able to use constructive criticism is essential. Suggestions, feedbacks and discussion must involve all relevant stakeholders and occur for every stage of policy experimentation and reform. This can only be achieved if a culture of open discussion prevails, with the objective, not of laying blame, but of discovering novel solutions that can be used to steer the system towards better outcomes. While governing complex education systems poses many challenges, educational systems must build risk-taking into their policy-making in order to continue to innovate and change.

Notes

1. Note that this is especially problematic if stakeholders have strong prior beliefs about the effectiveness or ineffectiveness of the policy under investigation.
2. The process is analysed in more detail in the case study by Edmondson, Roberto and Tucker (2002).

References

Brown, A.L. (1992), "Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings", *The Journal of the Learning Sciences,* Vol. 2/2, pp. 141-178.

Brown, L. and S.P. Osborne (2013), "Risk and innovation", *Public Management Review,* Vol. 15/2, pp. 186-208.

Burns, T. and K. Weatherby (2014), "Developing creativity and innovation in teaching", *Licensed to Create: Ten Essays on Improving Teacher Quality,* Royal Society for the Arts, United Kingdom, www.thersa.org/discover/publications-and-articles/reports/licensed-to-create/.

Campbell, D.T. (1969), "Reforms as experiments", *American Psychologist,* Vol. 24, pp. 409-429.

Cannon, M.D. and A.C. Edmondson (2005), "Failing to learn and learning to fail (intelligently): How great organisations put failure to work to innovate and improve", *Long Range Planning,* Vol. 38, pp. 299-319.

Chingos, M.M. and G.J. Whitehurst (2011), *Class Size: What Research Says and What it Means for State Policy*, Brookings Institution, Brown Center on Education Policy, Washington, D.C.

Edmondson, A., M.A. Roberto, and A. Tucker (2002), *Children's Hospital and Clinics,* Harvard Business School Case.

Fryer, R.G. (2011), "Financial incentives and student achievement: Evidence from randomised trials", *The Quarterly Journal of Economics,* Vol. 126, pp. 1755-1798.

Fullan, M. (2009), "The large-scale reform comes of age", *Journal of Educational Change,* Vol. 1/1, pp. 5-28.

Grossman, D. (2014), "Secret Google lab 'rewards staff for failure", BBC News, www.bbc.com/news/technology-25880738 (accessed 25 January 2014).

Hood, C. (2002), "The risk game and the blame game", *Government and Opposition*, Vol. 37/1, pp. 15-37.

Jensen, B., et al. (2016), *Beyond PD: Teacher Professional Learning in High-Performing Systems*, National Center on Education and the Economy, Washington, D.C. www.ncee.org/wp-content/uploads/2015/08/BeyondPDWeb.pdf

Kanouse, D.E. and L. Hanson, (1972), "Negativity in evaluations", in E.E. Jones, D.E. Kanouse, S. Valins, H.H. Kelley, R.E. Nisbett, and B. Weiner (eds.), *Attribution: Perceiving the Causes of Behaviour,* General Learning Press, Morristown, New Jersey, pp. 47-62.

Knight, F.H. (1921), *Risk, Uncertainty, and Profit*, Hart, Schaffner & Marx; Houghton Mifflin Company, Boston, MA.

Lau, R.R. (1985), "Two explanations for negativity effects in political behavior", *American Journal of Political Science*, Vol 29/1, pp. 119-138.

OECD (2015), *Schooling Redesigned: Towards Innovative Learning Systems*, Educational Research and Innovation, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264245914-en.

Osborne, S and S-S. Flemig (2015), "Conceptualising risk and social innovation: An integrated framework for risk governance", *Society and Economy*, Vol. 37/2, pp. 165-182.

Popper, K.R. (1957), *The Poverty of Historicism,* Routledge, London.

Renn, O. (2008), *Risk Governance: Coping with Uncertainty in a Complex World,* Earthscan, London.

Schwerdt, G. et al. (2011), *Effects of Adult Education Vouchers on the Labor Market: Evidence from a Ranzomized Field Experiment,* IZA, Bonn.

Snyder, S. (2013), "The Simple, the complicated, and the complex: Educational reform through the lens of complexity theory", *OECD Education Working Papers*, No. 96, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k3txnpt1lnr-en

Weaver, R. K. (1986), "The politics of blame avoidance", *Journal of Public Policy*, Vol. 6/4, pp. 371-98.

Weaver, R. K. (1988), *Automatic Government*, Brookings Institution, Washington D.C.

Chapter 11.

Enhancing effective education governance

Tracey Burns, *Centre for Educational Research and Innovation, OECD*

Lucie Cerna, *Skills Beyond Schools, OECD*

Effective multi-level governance of complex education systems is a policy priority. As educational systems have decentralised, countries are increasingly looking for ways to balance responsiveness to local diversity with national attainment goals. The first part of this chapter explores the importance of trust for the governance of complex systems and highlights its interaction with the main themes of this volume. It shows that trust is indispensable for change and reform but also raises important questions about the right levels of trust for the governance of educational systems. The second part of the chapter suggests a way forward. It summarises the main findings on governance systems that emerge from this volume, focusing in particular on issues of complexity, accountability, capacity building and strategic thinking. It then ends with a look at the key elements of modern educational governance.

Introduction

Governing multi-level education systems effectively requires governance models that balance responsiveness to local diversity with the ability to ensure national objectives. This is a delicate equilibrium, one that is difficult to achieve given the complexity of the education system in many OECD countries. As a result, governance issues have moved up on the political and policy agendas, and countries are increasingly looking for examples of good practice and models that they can adapt to their own needs.

This volume set out to address some of the key challenges involved in governing modern education systems, looking specifically at complexity, accountability, capacity building and strategic thinking. Yet as raised in the Introduction, one element has not yet been mentioned, although it is the glue that holds it all together. That element is trust.

Trust

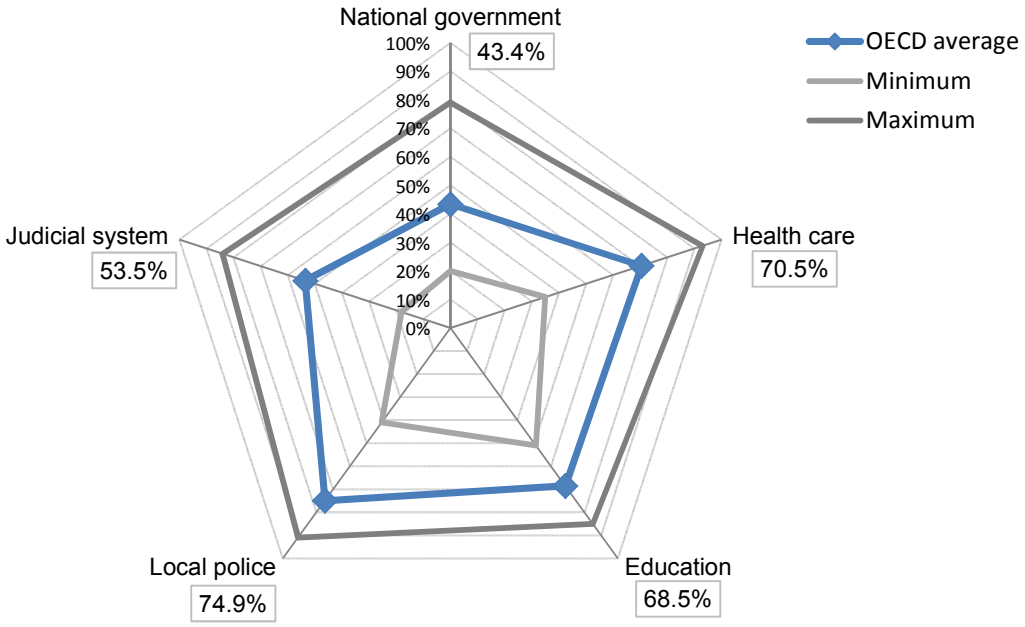
Trust impacts the governance of education systems in a number of ways. It enables stakeholders to take risks, facilitates interactions and co-operation, and reduces the need for control and monitoring (Levi, 1998; Van Maele, Forsyth and Van Houtte, 2014). Trust in the system impacts not only the functioning of the system, but also the actions of individual actors in the system. For example it affects:

- Consensus building across multiple stakeholders and different levels of government.
- The functioning, status, and professionalisation of teachers and school leaders.
- The educational planning of students and their parents.

Trust therefore offers flexibility to stakeholders to propose and implement innovative changes and reforms. It allows engaging parents, student and communities as active partners. Other factors such as high levels of professionalism and attractiveness of teaching depend on it. In short, although often neglected, high levels of trust enable smoother system functioning and facilitate the governance process. However, there is one caveat: too much trust can have a downside. Mistrust and distrust can also be appropriate responses in some circumstances, and blind trust in individuals and institutions can be abused. Determining an optimal level of trust is thus necessary, but as it will change over time and over contexts, it can be a difficult task (e.g. Wicks, Berman and Jones, 1999).

How do levels of trust in education compare with other sectors? Figure 11.1 compares the OECD average confidence levels for health care, education, local police, the judicial system and national government. In 2015, confidence was the highest in health care and the local police (70.5% and 74.9% respectively), followed by education (68.5%). People have more trust in concrete public services than in the abstract notion of national government.

Figure 11.1 Confidence in different segments of government, OECD average (2015)



Source: Gallup World Poll.

What does this mean for education, and for education governance more specifically? Education has a fundamental role in the development of trust and the cognitive abilities and social skills needed to understand and interpret others' behaviour (Borgonovi and Burns, 2015). It also plays a key role in effective governance of systems and system functioning more generally. The following sections will briefly examine how trust interacts with the main themes of this volume: complexity, accountability, working with stakeholders and networks, and new paradigms in governance such as experimentalism and risk-taking.

Complexity and trust

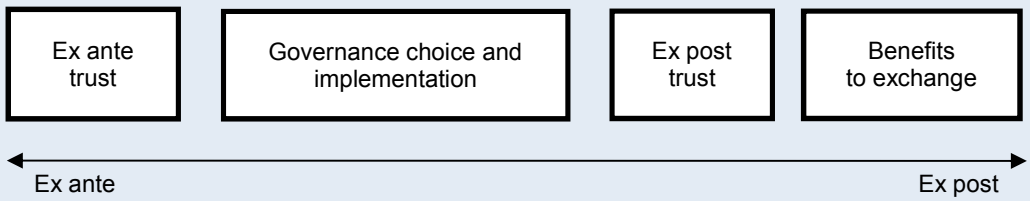
Many of the chapters in this volume highlight the complexity of education systems (see Chapter 2) and propose different ways of dealing with it, for instance, through policy experimentation (Chapter 8), risk-taking (Chapter 10), governance networks (Chapter 3), horizontal accountability (Chapter 5) or soft governance (Chapter 6). In addition to complexity, trust plays an important role in all these chapters, Nonetheless, evidence on the direction of the relationship between trust and complexity has been rather mixed (see Cerna, 2014a for a discussion of the literature). It is not clear whether there is a greater need for trust in complex governance systems than in simple, linear systems. Higher trust levels could be necessary in complex systems due to the number of stakeholders involved, who interact in non-linear ways. However, the complexity of governance systems might reduce trust levels as reliance on complex governance itself may signal a lack of trust to partners. Box 11.1 explores the relationship between trust and complexity in more detail.

Box 11.1. The relationship between trust and complexity

Puranam and Vanneste (2009) argue that there are three types of relationships between trust and governance complexity, which can coexist and can also play out at different points in time:

1. There is the relationship between ex ante levels of trust and the choice of governance mechanisms (Gulati, 1995). For instance, trust developed throughout repeated interactions between partners can influence their choice of governance structure in future alliances with each other. As Gulati (1995) argues, prior trust may allow actors to enter partnerships which may otherwise have been deemed impossible.
2. Complementarity can play out, which is the leveraging effect of a given level of trust on the relationship between governance and exchange performance (Poppo and Zenger, 2002). The value to increasing governance complexity is greater in the presence of trust than in its absence (Puranam and Vanneste, 2009).
3. Crowding out – direct and indirect – can take place, which is the influence of governance mechanisms on suppressing trust (Malhotra and Murnighan, 2002). In the case of a direct crowding out, the level of trust in the relationship can be eroded when partners rely on complex governance as this may signal a lack of trust to partners (Puranam and Vanneste, 2009: 15). In indirect crowding out, an introduction of formal governance mechanisms can hinder the process of trust formation because trustworthy behaviour can be attributed more to the governance mechanisms than to the partner (Malhotra and Murnighan, 2002).

The figure below illustrates these three kinds of governance relationships across several time periods. In the first period, the decision-maker chooses, at least in theory, an optimal level of governance complexity for the relationship, given an ex ante trust level, and implements it with some setup costs. In the second period, which starts after the implementation of the governance structure, the crowding out effect operates and the level of ex post control is determined. In the third period, trust and governance complexity act jointly and create benefits to the exchange, such as a smoother adaptation and the provision of safeguards. The decision-maker anticipates what is likely to happen in the second and third periods, and thus the first period decision on governance complexity accounts for this. In a next step, it would be important to examine the model and expectations in a variety of policy areas, such as education governance, in order to determine how it applies to real world situations.



Source : Puranam and Vanneste, 2009.

The chapters present various tensions arising from complexity. For instance, a combination of greater complexity in governance systems on the one hand and higher demands for accountability on the other hand create challenges for governments and stakeholders alike (Snyder, 2013). Trust can play a key role in reconciling these at times opposing goals by bringing benefits to the relationship and increasing co-operation between stakeholders, rather than creating conflict.

Accountability and trust

Accountability as a key ingredient of governance is closely related to information and communication, and trust; these are all among the soft factors in the improvement of governance (Chapter 6). While some forms of accountability can contribute to trust, one risk is that accountability can be a source rather than a remedy for distrust (O'Neill, 2005, see also Chapter 4). Some argue that accountability is an alternative to trust, since efforts to strengthen it usually involve parallel efforts to weaken trust (Trow, 1996: 3). However, there are also examples where trust and accountability can be balanced (see Hopfenbeck et al., 2013).

Multiple stakeholder accountability seeks to strengthen both trust and accountability by including more stakeholders in accountability processes (see Chapter 5). For it to work, it is important to recognise different interests and needs among stakeholders, allow enough time to develop a trusting relationship and clarify roles and purposes such that all actors feel responsible (Hooge, Burns and Wilkoszewski, 2012). However, this is a nuanced and difficult task, and differently performing schools may need different accountability systems (see Hooge, Burns and Wilkoszewski, 2012). Contrary to common beliefs, both accountability and trust are possible in a system, though their levels have to be balanced carefully (Chapter 1, also Busemeyer and Vossiek, 2015).

Collaboration and capacity building and trust

For multiple stakeholder accountability to work, collaboration between stakeholders is key, which relies on trust (see Chapter 3). At the school level, collaboration and continuous peer support are needed for evidence informed practice (Chapter 7). Collaboration enables learning from close observations of knowledge exchange and teaching exchanges. Peer support between teachers speeds up the process of developing trust that enables unlearning of old assumptions and habits as well as the development of new understandings and practices.

Greater collaboration between actors may thus create more trust as partners have experience with each other over time and can witness the benevolence, reliability, competence, honesty and openness of their partners. Putnam (2000) refers to this accumulation of collective trust as social capital. The resources of social capital - the norms, networks and trust - can increase a society's efficiency in solving collective action problems (Putnam, 1993). Networks with closure (i.e. dense networks) are the source of social capital as they facilitate access to information and sanctions (Coleman, 1988). Thus trust is an important ingredient in collaboration between partners and in networks between various stakeholders. The latter will be analysed further in the next section.

Networks and "soft" modes of governance and trust

Networks are considered one mechanism to deal with complexity; as they can cut through complex hierarchies and generate new solutions to intractable and often challenging local problems (OECD, 2015). Nonetheless, networks can create their own challenges (see Chapter 3). The reason for this is that many actors with different perspectives are involved, and there are tensions between the more centralised structure of ministries and the more horizontal nature of networks in education systems.

Trust is important for networks to operate; especially where power resources are diffused among actors and interdependent relations are strongly present (Edelenbos and Klijn, 2007: 26). They function because people are willing to co-operate and sacrifice

short-term gains for the benefit of long-term co-operation. Networks are different in this sense from markets and hierarchies, where the need for trust is minimised through complex systems of incentives and rules (Cerna, 2014a). In networks, the complexity of decision making and multiplicity of actors require investments in forming and maintaining relations (Agranoff and Mcguire, 2003). Trust can stimulate that investment and the effort actors put in those relations. Some flexibility in governance arrangements is also important for developing soft governance mechanisms (see Chapter 6).

Communication and trust, as well as information, dialogue and capacity building are considered key soft factors in the improvement of governance (see Fullan, 2011). Soft policy making as a complement to hard policy making in a centralised environment depends strongly on soft elements such as flexible resources, self-organisation and mobilising commitment (Chapter 6). Policy-makers need to provide an environment with the necessary and sufficient conditions for the development of soft factors/ elements. For instance, enabling good communication and flow of information between stakeholders is equally important as providing sufficient autonomy and flexible resources to stakeholders.

Risk-taking, innovation and experimentation and trust

Besides networks and soft policy making, other mechanisms to respond to complex systems are risk-taking, innovation and experimentation (see Chapters 8, 9, and 10). Trust is not taking risk per se, but rather it is a willingness to take risk (Mayer, Davis and Schoorman, 1995, McEvily et al., 2003). Even though several authors have recognised the importance of risk to understanding trust (Coleman, 1990; Lewis and Weigert, 1985; Luhmann, 1988; March and Shapira, 1987), no consensus on their relationship exists. It is unclear whether risk is an antecedent to trust, is trust, or is an outcome of trust (Mayer, Davis and Schoorman, 1995).

Encouraging teachers to take risks calls for governance processes able to manage the risks of failure (Chapter 7). From a governance perspective, policy makers need to recognise failure as an inherent part of policies and reforms, encourage constructive criticism and build the capacity and knowledge systems to learn from failure (see Chapter 10). This is also important for policy experimentation, which is another instrument of policy making in complex environments (see Chapter 8).

As discussed in Chapter 1, there is an inherent tension between accountability and innovation because careful accountability seeks to minimise risk and error, both of which are fundamental elements of the innovation process (Giddens, 1990; Reina and Reina, 2006). Yet countries are under strong internal and external pressure to strengthen their accountability systems (see Fullan, 2011), while at the same time encouraging innovation. Trust might be able to resolve this strong tension between accountability and innovation (Cerna, 2014b). Hence, establishing and maintaining trust – both in the classroom and the education system as a whole – is a crucial step in encouraging innovation and risk-taking.

A gap in our knowledge: The breakdown of trust

Clearly trust is important for governing complex education systems and enabling reform and change. In the presence of trust within a network, communication flow is improved and a sense of psychological safety among members is increased (Edmondson, 2004). But what to do when trust is broken? A common misconception is that once non-co-operation or untrustworthy behaviour is observed, a return to the co-operative or trust solution is not possible. However, there is some empirical evidence that trust in different

forms can be rebuilt (Jonker et al., 2004). For example, trust could be enhanced by greater communication and transparency between different stakeholders (Carless, 2009).

Rebuilding trust, however, is a lengthy and difficult process. The involvement of societal stakeholders and private actors in networks (a flexible form of governance) can generate more information and knowledge, which can be used to develop better tailored solutions for problems (Edelenbos and Klijn, 2006; Sorensen and Torfing, 2007) and potentially also build trust. There is ongoing work in our research centre on rebuilding trust, which will be explored in an upcoming publication (Volume II in this series [Burns et al., forthcoming]). Besides showcasing good practices in rebuilding trust, Volume II will provide empirical examples of innovative ways in education systems dealing with complexity and finding the right balance between appropriate levels of accountability, capacity building and strategic thinking in each context.

Taking stock: Main findings and a way forward

Main findings

Trust is thus interwoven into all the discussions of the volume. It is an essential element of educational governance and is required for good system functioning. Trust thus also underpins the main findings presented below.

From this volume, and the work of the OECD's Governing Complex Education Systems project more generally, a number of main findings have emerged. They are presented here as a series of observations:

- The first observation is that *there are no magic solutions, no one-size-fits-all recipe that can be rolled out to guarantee success.*

Modern education governance must be able to juggle dynamism and complexity at the same time as it steers a clear course towards established goals. And it must do this as efficiently as possible, with limited financial resources and increasing demands. Aligning multi-level systems and engaging with a diverse set of actors, including students and parents, is a challenging task. Education is a field with strong a priori beliefs, strongly tied both to our identities and our experiences. Not only do we expect education to deliver the kinds of citizens we desire, but everyone has taken part in education in some form or another. In doing so they have often formed strong personal opinions about what appears to work, and what does not, and these opinions may not be aligned with research findings. This makes dialogue and discussion essential, but potentially prone to subjective arguments and reasoning. Harnessing the power of communities and parental voice, while essential, can also be a delicate task, as explained below.

- The second observation is that *effective governance works through building capacity, open dialogue, and stakeholder involvement.*

While seeking consensus and common solutions is an established part of any governance process, the modern world has added a new twist: New technologies provide the opportunity to reach out to a broader set of actors and to take their views and concerns into account. However, new technologies also come with new challenges. For example, the opportunity for almost instant feedback can become a challenge, in that parents are not inclined to wait and see what is effective; they expect the best education for their children now. The danger here is that expectations tend to rise faster than performance, and there is a temptation for elected officials to operate in the short-term even though research has demonstrated that the effects of a reform can take a significant amount of

time to bear fruit (see Chapter 1 for a full discussion). Despite this, there is no going back: These kinds of horizontal accountability measures are essential compliments to the more traditional vertical accountability that has historically been used in most education systems. It is through the involvement and engagement of a diverse group of actors that educational governance will be able to continue to evolve along with our societies and schools. There is thus a need for mechanisms to include all stakeholders and voices (not only the most vocal or technologically savvy) in the governance process and designing ways to strengthen participatory governance mechanisms. This will also require working with less active or less confident stakeholders to build capacity and empowerment to enable them to take part in the process.

- The third observation is that *a whole of systems approach is essential.*

In complex systems nothing can be done in isolation, as it is the relationships between the parts that are essential. Simple solutions to complex problems are ultimately ineffective. Education systems must resolve tensions between potentially conflicting forces such as accountability and trust, innovation and risk-avoidance, and consensus building and making difficult choices. Finding the right balance (or, perhaps more accurately, the right combination of mutually reinforcing dynamics that are designed to strengthen both accountability and trust (for example)), will depend on the context and history of the system, as well as the ambitions and expectations for its future. As already stated, trust can play a key role in reconciling these at times opposing goals by bringing benefits to the relationship and increasing co-operation between stakeholders, rather than creating conflict.

But trust is not enough: finding the right balance will depend on the context of the system, as well as the ambitions and expectations for its future. A whole of systems approach works to align roles and responsibilities across the system, improving efficiency as well as reducing potential overlap or conflict. This approach thus necessarily includes a holistic long-term vision and strategy. In order to make this possible, a strong knowledge system that builds on rich and nuanced data that are easily understandable is required. This system combines descriptive system data (on achievement, graduation, etc.) with research findings that can determine whether something is working, and why. It also includes the wealth of practitioner knowledge available, both formalised and informal. In developing a whole of systems approach, the key is to knowing what to use, when, and why (Fazekas and Burns, 2012).

- This observation is related to the fourth observation, which is that *there are systemic weaknesses in capacity which contribute to today's governance challenges.*

A key element of successful governance is ensuring that stakeholders have sufficient capacity to assume their roles and deliver on their responsibilities. In particular, they need adequate knowledge of educational policy goals and consequences, the ownership and willingness to make the change, and the tools to implement a reform as planned. Many of these elements emerge from participatory governance processes and open dialogue, which serve to strengthen the legitimacy and ownership of the goals and process. Without these, the best policy reform risks being derailed at the level where it counts most: the classroom.

However, even with appropriate knowledge and ownership changes in roles and responsibilities generally also require explicit capacity building. For example, as schools become more autonomous headmasters have been given new roles and powers regarding

planning, budget, and staff. In some systems these are entirely new responsibilities and they must be given the support they need to grow into them. Another example is the use of data: in all systems there is more data available from system-level indicators, evaluations, and test scores, and capacity must be developed in order to use and interpret the data correctly. Note that even though the focus is often placed on the local level in discussions of capacity, systemic weaknesses have been observed on every level of governance, especially in the ability to use data and research evidence for policy-making.

In addition to capacity issues for new roles as mentioned above, another major issue is weak strategic thinking. As spelled out above, a whole of systems approach is essential, and this includes strategic thinking for the longer term. While there are a number of political pressures (including time) that work against the ability to plan and steer systems with a longer term system vision, a lack of capacity for strategic thinking, especially in smaller communities, continues to be a system challenge in many countries.

- Despite these weaknesses, the fifth observation is that even in decentralised systems, *the national or state level remains very important* in triggering and steering education reform.

The national (or state in some systems) level most often provides the leadership for the system-wide vision needed to enable effective delivery of reform as well as equitable access and outcomes for students. It can also be instrumental in developing clear guidelines and goals, and providing feedback on the progress on those goals, the building blocks of any successful governance and reform process. This level can and should provide the leadership and co-ordination for the development of the whole of system approach. This approach should not be developed in a top-down or isolated manner, but rather in partnership with a broad set of stakeholders (see also the second, third, and fourth observations above).

- Last but certainly not least, the sixth observation is that *there is a need to develop key principles for system governance* (not just agreement on where to go, but *how to get there*).

Examples of goals include reducing the drop-out rate and improving student attainment. Examples of key principles underlying the governance and decision-making used to achieve those goals would be having a system that is open, inclusive, positive, and evidence-informed. This then allows for all the elements set out in the previous observations, that is:

  - Stakeholder involvement and ownership of agreed goals and principles.
  - A whole of system vision that keeps the focus on processes, and does not get mired in discussing structures.
  - Alignment of roles and responsibilities across the system, as well as a way to address any potential conflicts or overlap.
  - The ability to identify needs and develop capacity in a realistic and timely manner, based on the system vision and informed by research evidence.
  - A flexible and adaptive education system that can react to change and unexpected events by relying on its processes. This is essential in the governance of complex systems.

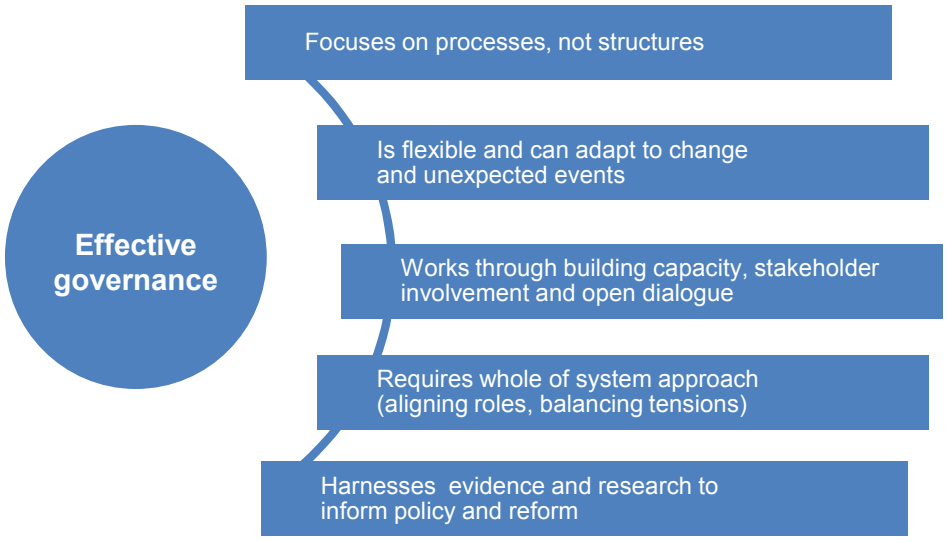
A way forward

This volume began with the description of a challenge: governing multi-level education systems effectively requires governance models that balance responsiveness to local diversity with the ability to ensure national objectives. This is a delicate equilibrium, one that is difficult to achieve given the complexity of the education system in many OECD countries. As a result, governance issues have moved up political and policy agendas and countries are increasingly looking for models that they can adapt to their own needs.

Yet the issue is not simple. Work on complexity theory reveals that a significant degree of complexity in a system – whether an education system or a school – leads to emergent properties beyond those predictable from initial conditions. The traditional policy cycle, which tends to frame planning and policy choices in a linear, reductionist manner, is no longer adequate. Modern education governance must be able to be flexible and adaptive at the same time as it steers a clear course towards established goals. And it must do this as efficiently as possible, with limited financial resources, and also within the confines and pressures of time-sensitive political cycles.

So what are the elements of effective modern governance systems? Based on the observations above, we propose a series of elements that keep the focus on process, allow systems to adapt and respond to complexity, and build on dialogue and participation of multiple actors. They keep knowledge and evidence at the core while at the same time supporting a system-wide vision of education and progress, as laid out in the figure below:

Figure 11.2 Elements of effective governance



Creating the open, dynamic and strategic governance systems necessary for governing complex systems is not an easy one. This volume challenges our traditional concepts of education governance through work on complexity, change/reform and new modes of collaborative networks and decision-making. In doing so it sets the agenda for thinking about inclusive, adaptable and flexible accountability and governance, necessary for governing complex systems in today's global world.

References

Agranoff, R. and M. McGuire (2003), *Collaborative Public Management: New Strategies for Local Governments*, Georgetown University Press, Washington.

Busemeyer, M. and Vossiek, J. (2015), "Reforming education governance through local capacity-building: a case study of the 'Learning Locally' programme in Germany", *OECD Education Working Papers*, No. 113, OECD Publishing, http://dx.doi.org/10.1787/5js6bhl2mxjg-en.

Carless, D. (2009), "Trust, distrust and their impact on assessment reform", *Assessment and Evaluation in Higher Education*, Vol. 34/1, pp. 79-89.

Borgonovi, F. and T. Burns (2015), "The educational roots of trust", *OECD Education Working Papers*, No. 119, OECD Publishing, Paris, http://dx.doi.org/10.1787/5js1kv85dfvd-en.

Burns, T., L. Cerna, M. Fuster and F. Köster (forthcoming), *Governance in Action: A Synthesis of Case Studies*, OECD Publishing, Paris.

Carless, D. (2009), "Trust, distrust and their impact on assessment reform", *Assessment and Evaluation in Higher Education*, Vol. 34/1, pp. 79-89.

Cerna, L. (2014a), "Trust: what it is and why it matters for governance and education", *OECD Education Working Papers*, No. 108, OECD Publishing, Paris, http://dx.doi.org/10.1787/5jxswcg0t6wl-en.

Cerna, L. (2014b), "Governance of innovation in education", *Education Sciences*, Vol. 2014/4, pp. 5-21.

Coleman, J. (1988), "Social capital in the creation of human capital", *American Journal of Sociology*, Vol. 94/suppl, pp. 95-120.

Coleman, J. (1990), *Foundations of Social Theory*, Harvard University Press, Cambridge.

Edelenbos, J. and E.H. Klijn (2006), "Managing stakeholder involvement in decision-making: a comparative analysis of six interactive processes in the Netherlands", *Journal of Public Administration Research and Theory*, Vol. 16/3, pp. 417-446.

Edmondson, A. (2004), "Psychological safety, trust and learning in organisations: A group-level lens", in R. Kramer and K. Cook (eds.), *Trust and Distrust in Organisations: Dilemmas and Approaches*, Russell Sage, New York, pp. 239-272.

Fazekas, M. and T. Burns (2012), "Exploring the Complex Interaction Between Governance and Knowledge in Education", *OECD Education Working Papers*, No. 67, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k9flcx2l340-en.

Fullan, M. (2011), "Choosing the wrong drivers for whole system reform", *Centre for Strategic Education Seminar Series Paper*, No. 204, Centre for Strategic Education, Melbourne.

Giddens, A. (1990), *The Consequences of Modernity*, Polity Press, Cambridge.

Gulati, R. (1995), "Does familiarity breed trust? The implications of repeated ties for contractual choice in alliances", *Academy of Management Journal,* Vol 38, pp. 85-112.

Hopfenbeck, T. et al. (2013), "Balancing trust and accountability? The assessment for learning programme in Norway: A Governing Complex Education Systems case study", *OECD Education Working Papers*, No. 97, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k3txnpqlsnn-en.

Hooge, E., T. Burns and H. Wilkoszewski (2012), "Looking beyond the numbers: Stakeholders and multiple school accountability", *OECD Education Working Papers*, No. 85, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k91dl7ct6q6-en.

Jonker, C. et al. (2004), "Human experiments in trust dynamics", *Lecture notes in Computer Science*, Vol. 2995, Springer, Berlin, pp. 206-220.

Levi, M. (1998), "A state of trust", in V. Braithwaite and M. Levi (eds.), *Trust and Governance*, Russell Sage, New York, pp. 77-101.

Lewis, J. and A. Weigert (1985), "Trust as a social reality", *Social Forces*, Vol. 63/4, pp. 967-985.

Luhmann, N. (1988), "Familiarity, confidence, trust: problems and alternatives", in D. Gambetta (ed.), *Trust: Making and Breaking Cooperative Relations*, Basil Blackwell, New York, pp. 94-107.

Malhotra, D. and J. Murnighan (2002), "The effects of contracts on interpersonal trust", *Administrative Science Quarterly*, Vol. 47/3, pp. 534-559.

March, J. and Z. Shapira (1987), "Managerial perspective on risk and risk-taking", *Management Science*, Vol. 33/11, pp. 1404-1418.

Mayer, R., J. Davis and F. Schoorman (1995), "An integrative model of organisational trust", *Academy of Management Review*, Vol. 20/3, pp. 709-734.

McEvily, B., V. Perrone and A. Zaheer (2003), "Trust as an organising principle", *Organisation Science*, Vol. 14/1, pp. 91-103.

O'Neill, O. (2005), "Justice, trust and accountability", in O. Neumaier, C. Sedmak and M. Zichy (eds.), *Gerechtigkeit: Auf der Suche nach einem Gleichgewicht* [Justice: Looking for a Balance], Ontos Verlag, Frankfurt, pp. 33-55.

OECD (2015), *Schooling Redesigned: Towards Innovative Learning Systems*, Educational Research and Innovation, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264245914-en.

Poppo, L. and T. Zenger (2002), "Do formal contracts and relational governance function as substitutes or complements?", *Strategic Management Journal*, Vol. 23/8, pp. 707-725.

Puranam, P. and B. Vanneste (2009), "Trust and governance: untangling a tangled web", *Academy of Management Review*, Vol. 34/1, pp. 11-31.

Putman, R. (1993), *Making Democracy Work*, Princeton University Press, Princeton.

Putnam, R. (2000), *Bowling Alone: The Collapse and Revival of American Community*, Simon and Schuster, New York.

Reina, D. and M. Reina (2006), *Trust and Betrayal in the Workplace: Building Effective Relationships in Your Organisation*, Berrett-Koehler Publishers, San Francisco.

Snyder, S. (2013), "The simple, the complicated, and the complex: Educational reform through the lens of complexity theory", *OECD Education Working Papers*, No. 96, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k3txnpt1lnr-en.

Sorensen, E. and J. Torfing (eds.) (2007), *Theories of Democratic Network Governance*, Edward Elgar, Cheltenham.

Trow, M. (1996), "Trust, markets and accountability in higher education: a comparative perspective", *Higher Education Policy*, Vol. 9/4, pp. 309-324.

Van Maele, D., P. Forsyth and M. Van Houtte (2014), *Trust and School Life: The Role of Trust for Learning, Teaching, Leading and Bridging*, Springer, Dordrecht.

Wicks, A., S. Berman and T. Jones (1999), "The structure of optimal trust: moral and strategic implications", *Academy of Management Review*, Vol. 24/1, pp. 99-116.

Contributors

Patrick Blanchenay

Patrick Blanchenay is a post-doctoral fellow at the University of Toronto. Prior to this position he was an economist in the OECD's Directorate for Science, Technology, and Innovation. He holds a PhD in Economics at the London School of Economics with a thesis on skill accumulation in agglomerations. Patrick Blanchenay further holds an MRes in Political Theory from Sciences Po Paris and an MSc in Management from HEC Paris.

Lex Borghans

Lex Borghans is professor in economics of education at the Department of Economics of the Maastricht University School of Business and Economics. His research covers educational choice, the development of cognitive and non-cognitive skills in all stages of the career, testing, and the economics of the curriculum. Furthermore, he has a special interest in the role of experiments in improving education.

Tracey Burns

Tracey Burns heads the Governing Complex Education Systems project in the OECD Centre for Educational Research and Innovation (CERI). She also directs the Centre's work on Trends Shaping Education. She holds a Master of Arts and PhD in Psychology from Northeastern University, USA. Previous to her current work Ms Burns worked on social determinants of health and on education and social inclusion issues at both the OECD and in Vancouver, Canada.

Lucie Cerna

Lucie Cerna is an analyst in the National Skills Strategy project at the OECD Directorate for Education and Skills. She holds a DPhil in Politics from the University of Oxford, where she focused on the governance of skilled immigration policy. Prior to coming to the OECD, she was a Lecturer in Politics at Merton College, Oxford, and an Assistant Professor in Global Challenges at Leiden University, the Netherlands. Ms Cerna holds an MSc in European Politics and Governance from the London School of Economics, and an MBA from Saarland University, Germany.

Philippa Cordingley

Philippa is Chief Executive of CUREE and an expert in evaluation, research and its use and effective Continuing Professional Development and Learning (CPDL). Ms Cordingley has led the CUREE team in a range of research, evaluation and development projects including a range of large scale evaluations. She has advised

numerous governments on topics such as standards, evidence information policy and practice and knowledge mobilisation. Ms Cordingley leads CUREE's work in designing and supporting 7 randomised controlled trials across 780 schools for the National College of Teaching and Leadership (NCTL).

Edith Hooge

Edith H. Hooge is full professor "Boards and Governance in Education" at TIAS School for Business and Society in the Netherlands. Her research activities revolve around governance and management in education systems and organisations. She teaches in the TIAS programmes for professionalisation of (non-)executive board members, and in the executive Masters for Public Management and for Management in Education. Ms Hooge is the director of TIAS GovernanceLAB She regularly presides in monitoring committees of governance codes and advises boards in different public sectors. In 1998 she completed her dissertation about "the policymaking of schools against the background of deregulation policies and increased autonomy" at the University of Amsterdam.

Florian Köster

Florian Köster is an analyst at the OECD in CERI's Governing Complex Education Systems project. He holds a MA in Government (specialising in Public Policy and Comparative Politics) from the University of Konstanz, Germany, and a MRes in Political Science from Universitat Pompeu Fabra, Spain.

Lorenz Lassnigg

Lorenz Lassnigg works as a senior researcher and head of the research group EQUI – Employment-Qualification-Innovation at the IHS-Institute for Advanced Studies in Vienna, Austria. Research interests are education policy and life-long learning in a wider institutional and historical perspective. Recent studies concern the governance and financing of Austrian schools and the understanding of the development of vocational education, in particular the apprenticeship system. A comparative analysis of the Survey of Adult Skills (a product of the OECD Programme for the International Assessment of Adult Competencies, or PIAAC) data has tried to trace the distribution of competences to the structures and reform dynamics of selected countries.

Mark Mason

Mark Mason is Professor in Comparative and International Education and Development at the Hong Kong (China) Institute of Education. He works from a disciplinary background in philosophy, social theory and education studies. Mr Mason has been the editor of a range of academic journals and monograph series, including the International Journal of Educational Development and the CERC Studies in Comparative Education series. He is a Past President of the Comparative Education Society of Hong Kong (China) and a former Director of the Comparative Education Research Centre at the University of Hong Kong (China). A former Fulbright Scholar, Mr Mason holds a doctorate from Columbia University.

William C. Smith

William C. Smith is a Senior Associate with RESULTS Educational Fund based in Washington, D.C. (United States). Prior to this position he completed his PhD in Educational Theory and Policy and Comparative International Education at The Pennsylvania State University and was a Thomas J. Alexander Fellow at the Organization for Economic Co-Operation and Development (OECD). His research addressing education's role in economic and social development has resulted in over 15 publications and includes an edited volume in the Oxford Studies in Comparative Education Series titled "The Global Testing Culture: Shaping Education Policy, Perspectives, and Practice" (2016).

Trudi Schils

Trudie Schils is an education economist working at the Department of Economics of the Maastricht University School of Business and Economics. Her research is inspired by the idea that education can only be improved if education practitioners, policymakers and researchers work together. Therefore, she cooperates with teachers, school boards, policy members and researchers from other disciplines. Her research focuses on the analysis of both student and school performance, for example the role and limitations of using tests to measure such performance as well as in the various determinants of performance.

Henno Theisens

Henno Theisens is a professor of Public Management at The Hague University for Applied Sciences. His work focusses on the effective governance of today's complex public systems, like education, health or public safety. He has extensive experience in the area of education on issues related to governance and long term strategy. Previously he was an analyst at OECD's Centre for Educational Research and Innovation and a Senior Research Associate at the Centre for Higher Education Policy Studies. Mr Theisens holds a PhD in Comparative Higher Education Policy Research from University of Twente and a MSc in European Politics and Policy from the London School of Economics and Political Science (LSE).

Inge de Wolf

Inge de Wolf is a professor of Education Systems at the department of Economics of Maastricht University in the Netherlands. She is initiator and manager of the "Academic Workplace in Education", a research network of Maastricht University, the Inspectorate of Education and several teachers and other professionals in education. Her research focuses on school improvement, education systems, student careers and evaluation. Inge also works as a coordinating inspector at the Inspectorate of Education. She is responsible for the yearly "State of Education in the Netherlands" and other system level evaluations of the inspectorate.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where governments work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The European Union takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

OECD PUBLISHING, 2, rue André-Pascal, 75775 PARIS CEDEX 16
(96 2016 03 1 P) ISBN 978-92-64-25535-7 – 2016

Educational Research and Innovation

Governing Education in a Complex World

What models of governance are effective in complex education systems? In all systems an increasing number of stakeholders are involved in designing, delivering and monitoring education. Like our societies, education systems are increasingly diverse regarding students, teachers and communities, as well as the values and identities we expect education to deliver. These trends have increased the complexity of education systems, leaving decision makers on all governance levels with the question of how to successfully manoeuvre in this highly dynamic policy area.

Governing Education in a Complex World addresses key challenges involved in governing modern education systems, looking specifically at complexity, accountability, capacity building and strategic thinking. The publication brings together research from the OECD Secretariat and invited chapters from international scholars to provide a state of the art analysis and a fresh perspective on some of the most challenging issues facing educational systems today.

Creating the open, dynamic and strategic governance systems necessary for governing complex systems is not easy. This volume challenges our traditional concepts of education governance through work on complexity, collaborative networks and decision-making. In doing so it sets the agenda for thinking about the inclusive and adaptable systems necessary for governing education in today's world. The volume will be a useful resource for those interested in education governance and complexity, particularly policy-makers, education leaders, teachers and the education research community.

Consult this publication on line at ***http://dx.doi.org/10.1787/9789264255364-en.***

This work is published on the OECD iLibrary, which gathers all OECD books, periodicals and statistical databases.
Visit ***www.oecd-ilibrary.org*** for more information.

OECD*publishing*
www.oecd.org/publishing

ISBN 978-92-64-25535-7
96 2016 03 1P