# Approaches to school evaluation in England since the Education Reform Act of 1988: rethinking the purpose, method and use of school evaluation based on expert interviews

**By Sevda Ozsezer-Kurnuc**

Thesis submitted for the degree of Doctor of Philosophy

University of East Anglia

School of Education and Lifelong Learning

**May 2022**

# ABSTRACT

Questions about the effectiveness of state schools in England in the 1980s, gave rise to debates on the quality of education, opening the system to external scrutiny. In this context, the Education Reform Act 1988 introduced a new approach to evaluating schools, underpinned by New Public Management, aimed at improving student performance in national tests and resting on four *pillars*: accountability, Ofsted inspections, school league tables and parental choice. This 'official' (government) approach to school evaluation has, to this day, remained controversial, attracting both support and critique. This study seeks to contribute to the ongoing debates about this approach by drawing on data collected through expert (elite) interviews. Fifteen educational experts working in the roles of education *policy actors*, *policy influencers*, and *school practitioners* were purposively selected and asked to assess the official approach to school evaluation and suggest alternatives. The interview data were analysed for similarities, differences and themes across the participant sample and interpreted through the lens of evaluation theory (Christie and Alkin, 2013; MacDonald, 1974, 1976; Patton, 2006, 2011, 2012). The views of expert participants point to a number of benefits as well as problems with the official approach to school evaluation. Their suggested alternatives range from minor changes to Ofsted's inspection methods to more radical alternatives, which include reframing inspections, reconfiguring accountability, and reconceptualising school evaluation.

This thesis concludes that the official approach to school evaluation in England introduced with the Education Reform Act 1988 does not appear to have worked as intended. It was designed as objective and evidence-based, but within a limited conception of evidence, based on the *pillars* of accountability, Ofsted inspection, school league tables and parental choice. To understand why this approach has not worked as intended, we need to examine the foundations of these four *pillars* in New Public Management and marketised education. In alignment with some of the experts who participated in this research, this thesis argues for an alternative approach to school evaluation which relies on trusted professionals, voice, capacity building, and collaboration. These alternative *pillars* rest on the foundations of democratic values, community involvement, and the evaluation of schools as complex, non-linear organisations that operate in specific contexts. The main purpose of complexity-based evaluation is collaborative learning, essential for improving both the single school and the whole education system.

## Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

# CONTENTS

# List of Tables

# List of Figures

# Acknowledgements

Victory is for those who can say '**Victory is mine**'. Success is for those who can begin saying 'I will succeed' and say 'I have succeeded' in the end.

Mustafa Kemal Atatürk

# List of Abbreviations and Acronyms

DfE          Department of Education

DfES        Department for Education and Skills

ERA         Education Reform Act

GCSE       General Certificate of Secondary Education

HMI         Her Majesty Inspectorate

LEA         Local Education Authority

MAT        Multi Academy Trust

NLE         National Leader for Education

NLG         National Leader for Governance

NPM        New Public Management

NSS         National Support School

Ofsted      Office For Standards in Education, Children's Services and Skills

RSCs        Regional Schools Commissioners

SATs        Standard Assessment Tests

SEF         School Self-evaluation Form

# CHAPTER 1: INTRODUCTION

## 1.1 Introduction, context, and contribution of this study

Evaluation is 'probably as old as the human race' (Shadish and Luellen, 2005: 183). Evaluative thinking and judgment could be seen as vital in helping humans to survive through the ages in a range of everyday activities, from building campfires to generating and accumulating knowledge of everyday situations. Whereas the evidence of 'formally organised evaluation' goes back hundreds of years, it was not until the 1970s that evaluation grew as a professional endeavour and the field of evaluation began to flourish (Shadish and Luellen, 2005: 183). This growth built on earlier developments in social science, which included the refinement and development of promising new applications of research methodologies to social problems, generously funded by central governments in the USA and the UK. Centrally-funded social programmes targeted a broad range of objectives, from the 1950s recovery from post-war economic problems to school curriculum reforms in the 1970s. Evaluation-related activities were also common in the private sector, with the aim of improving company profitability through accountants who monitored their financial activity, management consultants who worked to improve their operations, and research and development teams who designed and provided feedback regarding products (Hogan, 2007; Shadish and Luellen, 2005). These developments contributed to evaluation being established, by the 1990s, as a field of knowledge, with evaluation professionals being able to offer insight into 'which of several alternative actions tended to produce desirable results, improving operations, identifying needs to which programs could respond, justifying a program's budget, and creating support for a proposal or for continued funding of the program' (Shadish and Luellen, 2005: 185). However, by the late 1960s and 1970s, questions about the effectiveness of government investment in the public sector were also raised, in parallel with the rising conversations about accountability (Hogan, 2007; Picciotto, 2015), under the influence of a new approach to managing organisations in the private as well as the public sectors, New Public Management (NPM).

It is within this context that the Education Reform Act (ERA) became law in 1988. This thesis focuses on the official approach to school evaluation introduced with ERA and the subsequent policies for reforming education system in England (i.e., Education Act 2005, Schools White Paper 2010). The term 'official' refers in this thesis to an evolving approach to school evaluation introduced by the UK Government. The ERA of 1988 was a culmination of wide-sweeping changes to the evaluation of schools, introduced by successive governments in

England partly in response to the former Prime Minister James Callaghan's argument that rather than being 'secret gardens' schools should open themselves to external scrutiny (Chitty, 2014). These changes were influenced by the NPM, introduced to public services in the 1980s (Norris and Kushner, 2007), as well as legislation and policy papers such as ERA 1988 (Parliament, 1988); Education Act 2005 (Parliament, 2005) and Schools White Paper 2010 (DfE, 2010).

In the name of good governance, NPM focused on 'results and outcomes as opposed to activities and outputs' with a 'strong emphasis on accountability' (Levin-Rozalis et al., 2009: 191). It promoted the development of evaluative systems characterised by 'constant monitoring and the construction of targets and league tables for every public service' (Purdue, 2005: 123) with an intention to 'make performance transparent, but also intensifying control' (Norris and Kushner, 2007: 3). With the NPM, outcomes-based accountability, performance indicators, and targets as measures of productivity, traditionally deployed in the business sector, were introduced into education (Greene, 2009; Lane, 2000; Norris and Kushner, 2007). At the same time, 'evidence-based' educational evaluation was proposed as key to 'producing generalisable knowledge that will maximise educational resources (e.g., financial, human)' (Cousins and Ryan, 2009: 544). According to Norris and Kushner (2007: 5), the NPM found expression in educational policies through:

> local financial management and financial delegation, vouchers, and tax credits increasing parental choice and encouraging quasi-markets, national testing, and league tables, allowing schools to opt-out of local control, charter and trust schools, outsourcing schooling to for-profit and not-for-profit organisations, and public-private partnerships.

The 1988 ERA enshrined these in law, affirming what I will refer to in this thesis as the four *pillars* in the official (government) approach to school evaluation: accountability; a centralised inspection regime (as the precursor to the emergence, in 1992, of 'Ofsted' or the Office for Standards in Education); school league tables and customer (parental) choice in an education system that becomes a 'school market' (Cousins and Ryan, 2009; Gillard, 2011; Ryan and Feller, 2009). It is important to emphasise that accountability was not a new concept in 1988, but the Education Reform Act introduced a specific framing of the accountability of schools, to the central government, through centrally set school performance targets, as well as to the market, through parents as 'consumers' choosing the top-performing schools for their children. Where this thesis refers to 'accountability' as one of the four pillars of ERA 1988, it denotes this 'performative' approach introduced by ERA. However, as explored further in Chapters 6

and 7, 'accountability' is not a 'universal' concept with a uniform meaning. For example, Ranson's (2007) analysis points to a shift in the 1980s from professional accountability to accountability narrowly conceived through contract, corporate, performative and consumer terms. Norris and Kushner (2007) focused on the distinction between performance-based and outcome-based accountability. Wilkins' (2017) analysis pointed to a shift to 'centralised accountability' and its consequences for school governance. Similarly, some of the expert participants in this study problematised the notion of 'performative' accountability and suggested alternative conceptualisations (see Chapter 6, Section 6.3.2 and Chapter 7, Section 7.4.3).

It is also important to emphasise from the outset that, since 1988, further policies (e.g., Education Act 2005, Schools White Paper 2010) were introduced by successive governments but each preserved the basic system established in 1988. In other words, the four *pillars* remained fundamental in the official approach to school evaluation. Does this mean that the four *pillars* have worked well, or have they been taken for granted?

According to Norris and Kushner (2007: 9), tying up the systems of evaluation to 'rewards and punishments' may reduce openness and honesty in favour of carefully presented 'appearances.' Also, increasing demands for evidence-based accountability, as opposed to trusting professionals working in schools, has been a sign of the growth of low-trust accountability regimes. As modern evaluative mechanisms demand conformity with 'prescribed patterns of action' and depend on the calculability of results (Norris and Kushner, 2007: 7), some unintended consequences may appear such as fabricating performance and documents, decreasing creativity of schools, and performing 'on the day' for Ofsted inspectors (Ball, 2003; Ehren et al., 2016; Perryman, 2006, 2009).

Despite these negative consequences, the commitment to performance indicators in school evaluation has increased since the 1980s. As much as these *pillars* have many advocates (Barber, 2004; Gilbert, 2012; Matthews and Smith, 1995), they have also received much criticism (Ball, 2003; Brighouse and Waters, 2022; Brimblecombe et al., 1995; Gewirtz et al., 1994; Richards, 2015; Rosenthal, 2004; Torrance, 2011). Since the 1980s, therefore, approaches to school evaluation have been a controversial and important topic of debate and research. However, a search conducted on 20 February 2022 in ERIC (Educational Resources Information Center) revealed no peer-reviewed publications that investigate all these *pillars* altogether. For example, an advanced search using the keywords: 'school evaluation', '1988

Education Reform Act' and 'England/UK' referred to just 12 peer-reviewed articles published in academic journals. Their authors focused on investigating one of the *pillars* of school evaluation at a time (Barham, 1996; Hellawell, 1992; Hoskins, 2012; Pagett, 1996: Richmond, 1996; Wikeley and Hughes, 1995) or critically examining ERA 1988 through the conceptual lenses such as centralisation, professionalism, and leadership (Chychuk, 2015; Fisher, 2008; Gray, 2007; Hammersley-Fletcher, 2007; Hudson, 2007; Strain, 2009). These articles were mostly desk-based reviews rather than empirical studies and the majority were written before 2000.

This thesis, therefore, aims to contribute to knowledge and debates on school evaluation in England by developing a comprehensive assessment of the four *pillars*, their foundations in marketisation and New Public Management, as well as their overarching purpose framed as the improvement of the single school rather than the whole education system. It is also envisaged that rich interview data offered by educational experts may contribute new insights into alternatives to the official approach to school evaluation in England. A limited number of alternatives to the official approach have been suggested in the literature (Chapman, 2000; Matthews and Headon, 2015; Thrupp, 2005). An advanced search for 'school evaluation', 'alternative' and 'England/UK' in the ERIC database revealed no peer-reviewed publications about 'alternatives' to school evaluation in England. Six publications offered alternatives for student assessment (Alden, 2018; Daugherty et al., 2008; MacDonald, 2016; McCluskey, 2017; Richmond, 2017; Tan et al., 2021), one article offered an alternative for external accountability of Multi Academy Trusts (MATs) (Ehren and Godfrey, 2017), one article suggested alternatives to the performance-based accountability measures (Volante, 2015), three articles were about alternatives to the curriculum (Beck, 2012; Berry, 2009; Brundrett, 2015), and one article focused on alternatives to Ofsted grades and frameworks (Richards, 2015). Therefore, it is hoped that this thesis will contribute important new insights into alternatives or improvements to the official approach, based on findings from expert interviews presented in Chapters 6, 7 and 8, as well as evaluation theorists presented in Chapter 3.

## 1.2 Research aims, research questions and my motivation as a researcher

The aims of this research are as follows. Firstly, to investigate the evolving approach to school evaluation in England since the 1988 ERA. Secondly, to examine educational experts' views on the official approach. Thirdly, to consider the suggested alternatives to the current official

approach. The data collection method of expert (elite) interview was employed to capture the comprehensive perspectives of educational experts working in the education system in the roles of *policy actors*, *policy influencers*, and *school practitioners* as *policy implementers* or 'policy takers' (Gunter and Forrester, 2009) (see also Chapters 4 and 5).

This study has been guided by the following research questions:

1. What are educational experts' views on the official approach to school evaluation promoted in England since 1988?

2. Which alternatives to the official approach have been suggested by research participants?

3. How can these alternatives be assessed in the light of knowledge in the field of evaluation?

This topic is of personal importance to me as a student on a scholarship from the Turkish government and a prospective policymaker responsible for school evaluation and inspection in the Turkish education system. Enhancing my understanding of this topic would afford me an opportunity to understand the evolving approach to school evaluation in the English education system, as well as its underpinning methodologies. Whilst this thesis focuses on the English system, this research might help me to take insights from educational experts who have generously given me their time to contribute to improving the Turkish school evaluation system in the future.

## 1.3 Significance of the study

In addition to the contribution of this research discussed in Section 1.1 above, the significance of this study also lies in the research design itself, in the form of expert (elite) interviews. This thesis proposes a specific definition of an 'educational expert', based on professional knowledge and experience rather than power and status (Van Audenhove, 2017). It also reflects on the experience of a female international junior researcher preparing for expert (elite) interviews and developing strategies to cope with challenges of interviewing educational experts, both anticipated and unanticipated. The pandemic that we have all experienced has proved to us how problematic it is to evaluate schools in terms of accountability, Ofsted inspections, school league tables and parental choice. These four *pillars* may have fulfilled the function of driving school improvement before 2019. However, the pandemic introduced an unprecedented 'new normal' (e.g., full-time online schooling) and cancelled, for a year, the 'old normal' of standardised high stakes tests, as well as pausing, for a while, Ofsted inspection

visits. Whilst some experts who participated in this study put forth strong arguments for rethinking the official approach to school evaluation before the pandemic struck, recasting the four *pillars* of school evaluation may be even more important at the present time, as schools begin to embark on educational recovery from the pandemic. It is hoped that this thesis may contribute to the calls to move from the age of 'markets, centralisation, and managerialism' in order to renew 'hope, ambition, and collaborative partnership' (Brighouse and Waters, 2022: 1-23).

## 1.4 Structure of the thesis

Following this introductory chapter, Chapter 2 reviews the core school evaluation literature to address the first aim of this research and investigate the evolving approach to school evaluation in England since the 1988 Education Reform Act (ERA). To highlight the influences and decisions made before 1988, the chapter starts with a brief outline of approaches to school evaluation 'pre-1988'. The chapter proceeds with a more detailed discussion of ERA and its four *pillars:* accountability, Ofsted inspections, school league tables and parental choice. The review of the four *pillars* of school evaluation covers the perspectives of their 'advocates' and 'critics'. The chapter also outlines evolving approaches to self-evaluation, followed by a discussion of the methodological underpinnings and alternatives to the official approach suggested in the literature. The review of published empirical research and debates on the *pillars* supporting school evaluation in England presented in this chapter highlights that the foundations on which these *pillars* rest and the assumptions about how they drive school improvement may not be fit for purpose.

Chapter 3 develops a conceptual framework based on evaluation theory. It uses mainly Christie and Alkin's (2013) Evaluation Theory Tree, a metaphor which captures the development of the many branches of evaluation, as well as their roots in the 'foundational ideas' of *social accountability*, *social inquiry* and *epistemology*. The conceptual framework also uses the distinction between 'traditional' and 'developmental' approaches to evaluation developed by Patton (2006, 2011, 2012), discussing them in terms of their distinctive methodologies, differences in the use of evaluation findings, as well as contrasting perspectives on valuing (judging) and values. The chapter also focuses on democratic evaluation (MacDonald, 1974, 1976; Norris, 2015; Picciotto, 2015; Simons, 1987) as a distinctive approach developed in the UK. The discussion then moves to rethinking school evaluation in England. The conceptual framework draws on concepts from developmental evaluation and democratic evaluation that

are helpful in rethinking the purpose, method, use and valuing in school evaluation.

The methodology employed in this study is presented in Chapter 4. The chapter explains my epistemological and ontological stance, with my understanding of a given approach to school evaluation as a social construct. The chapter also discusses expert interviews and how this method of data collection enabled me to answer my research questions. I also discuss research ethics, the approach to the analysis of interview data, as well as issues related to the validity, reliability, and generalisability of this research.

Chapter 5 introduces educational experts who participated in this study and discusses three categories of education expertise developed by the participants, based on their roles in the education system: *policy actors*, *policy influencers* and *school practitioners* as *policy implementers* or 'policy takers' (Gunter and Forrester, 2009).

Chapter 6 and Chapter 7 present the findings from the interview data. Chapter 6 addresses research question 1 by presenting the views of educational experts on the official approach to school evaluation promoted in England since 1988. Chapter 7 answers research question 2 by presenting alternative approaches to the current official approach suggested by the participants. It is important to explain that these two chapters seek to foreground the views of educational experts, interspersing these with some references to published literature. The presentation of interview data in these chapters brings together emerging patterns and themes.

Chapter 8 discusses the key findings presented in Chapters 6 and 7 to address research question 3, by assessing the suggested alternatives in the light of knowledge in the field of evaluation. The discussion is informed by the conceptual framework of this research. 'Factual' and 'conceptual' findings (Trafford and Leshem, 2008: 133) are discussed and gathered together to recast the four *pillars* of the official approach as: trusted professionals, voice, capacity building, and collaboration. These alternative *pillars* rest on the foundations of democratic values, community involvement, and the evaluation of schools as complex, non-linear organisations that operate in specific contexts, aimed at collaborative learning and public accountability.

Chapter 9 concludes the thesis by considering the implications of these findings. The chapter summarises the key findings in relation to the research questions and presents the contribution of this study to knowledge. The chapter also considers the limitations of the study, suggesting topics for future research and shares my personal reflection on my research journey.

# CHAPTER 2: LITERATURE REVIEW

## 2.1 Introduction

There was a time with no league tables, no Ofsted, and no preoccupation with parental choice in the English education system. Educational improvement relied on the professional judgement of teachers and schools operated without external scrutiny, in what James Callaghan referred to as 'secret gardens' (Shaw, 2011, Chitty, 2014). The 1988 Education Reform Act (ERA) became a turning point in England, introducing school evaluation which has become increasingly centralised by successive governments to focus on what I refer to in my thesis as the official (government) approach. Driven by the principles of New Public Management (NPM) and marketisation, the official approach stems from a premise that the overarching aims of school evaluation are to do with (single) school improvement and public accountability, resting on the four *pillars* of school evaluation: accountability, Ofsted inspections, national testing ('high stakes tests') reported in school performance tables ('league tables'), and parental ('customer') choice. The issue of public accountability is particularly important because there appears to be a circular argument whereby accountability is construed as both a driver of school improvement (i.e., a *pillar* on which school improvement rests) and an overarching purpose of school evaluation. The term 'public accountability', as noted by Gilbert (2012), was introduced around 1992 as distinctive from 'local accountability' of schools pre-1992.

No successive government has departed from these fundamental four *pillars* of the ERA, although further policies have sought to modify some aspects of school evaluation, for example Education Act 2005 and Schools White Paper 2010. This chapter reviews published literature to explore these changes and the debates on the official approach to school evaluation in England since 1988. The chapter starts with a brief outline of approaches to school evaluation 'pre-1988' to highlight the influences and decisions made before 1988 (Section 2.2). The chapter then proceeds with a more detailed discussion of ERA and its four *pillars* (Section 2.3). Section 2.4 outlines evolving approaches to self-evaluation and is followed by a discussion of the methodological underpinnings of the official approach in Section 2.5. Section 2.6 focuses on alternatives to the official approach suggested in the literature. The review of published empirical research and debates on the *pillars* supporting school evaluation in England presented in this chapter highlights that the foundations on which these *pillars* rest and the assumptions about how they can drive school improvement may not be fit for purpose.

## 2.2 School Evaluation pre-1988

School evaluation pre-1988 was the official responsibility of both Her Majesty's Inspectorate (HMI) and local inspectorate. Both bodies originated in the 19th century in the need to 'supervise the expenditure of public money' and worked as complementary to each other (Plowden, 1967: 335). The oversight of the system depended mainly on individual HMI Inspector's judgement, particularly in terms of its efficiency (Ozga and Lawn, 2014). In the late 19th century, HMI inspectors visited elementary ('primary') schools annually to 'test' children's skills in reading, writing, and arithmetic, and schools were paid depending on whether children passed tests set by HMI inspectors (HMI, 1970; Shaw, 2011). These 'extensive powers' of HMIs over 'elementary' education first started to be devolved (HMI, 1970) with the establishment of Local Education Authorities (LEAs) in the 1902 Education Act. But it was through the 1944 Butler Education Act that LEAs' status was increased as they were given more responsibility to 'ensure that there was sufficient provision for the educational needs of pupils in their geographical area' (Shaw, 2011: 6). Importantly, main 'power' in education rested with the middle tier of local government, the role of the Secretary of State for Education was strictly limited to general policies with the purposes outlined in Butler's Education Act 1944. It was 'up to each LEA to decide whether to accept the minister's suggestions' (Brighouse and Waters, 2022: 13).

Individual teachers were able to decide what to teach (apart from religious education) and how to teach it. They had the power to change their practices if they wished without any structural constraints (Elliot, 1993; Shaw, 2011). Unsurprisingly, teachers were seen as professionals contributing to a 'better society' (Brighouse and Waters, 2022). Headteachers, in consultation with governors, had a role of controlling the school curriculum and resourcing (Shaw, 2011) and there was no expectation that the national government would control the curriculum (Brighouse and Waters, 2022). Educational improvement thus relied on the professional judgement of teachers and schools operated with little external scrutiny (Chitty, 2014; Shaw, 2011). In this context, self-evaluation by 'GRIDS', a form of self-evaluation using a set of criteria for schools to review their performance (MacBeath, 2005: 10), emerged in the 1970s in London schools, encouraged by their LEAs and supported by local advisers.

In the 1940s to 1960s, HMI inspected schools once every seven years to provide 'unbiased expert opinion' on performance (Ozga and Lawn, 2014: 9). They visited classrooms, 'not to judge a teacher' but 'to assess the quality of the work in progress'. HMI's advice was 'no more

than advice' and teachers were 'free to choose the course of action that appears to him the best' (HMI, 1970: 15-16). Therefore, HMI Inspectors' visits were directed towards 'informal and constructive discussion rather than inspection in the strict sense of the word' (Plowden, 1967: 384). They undertook regular full inspections, with teams of up to 15, as well as short inspections on aspects of schools (Elliot, 2012). The report on an inspection was 'comprehensive' and 'formal', covered 'all aspects of a school's life and work', and was for the Department of Education, the school staff, the governors, and the LEA (HMI, 1970: 24). HMI inspectors' remit was also to keep the ministers 'informed' about the progress of education in England (HMI, 1970) and advise them on the state of publicly funded education (Elliot, 2012: 1). By the 1970s, school inspection was also taken up by local authorities, which developed their own inspection and advisory services to support local schools (Ozga and Lawn, 2014). Whilst the local authority inspectorate possessed a more intimate knowledge of the work of schools in their area, HMI generally contributed from a wider view of educational practice throughout the country (Elliot, 2012; HMI, 1970).

HMI was reported to be 'respected' by schools due to their 'prestige, high reputation and expertise' which connected to the language they used to encourage good schools and 'gently rebuke' struggling schools and the discursive power they exercised by a 'deliberate eschewing of authority' (Elliot, 2012: 1; Kogan, 1970: 20, cited in Ozga and Lawn: 9). However, concerns about HMI were also highlighted, for example the 'progressive and often radical' approach of HMI (Bolton, 1998: 335, cited in Ozga and Lawn: 11). According to Dunford (2017: 5), the reputation and influence of HMI suffered in the 1950s as school reports were not published; but for Plowden (1967), the limited number of HMI Inspectors was the main reason for problems such as the low frequency of inspection visits. According to Brighouse and Waters (2022: 13), in his speech at Ruskin College in 1976, Prime Minister James Callaghan left the role of the inspectorate 'deliberately uncertain'. There was also a growing dissatisfaction of the Government with HMI reports that 'seemed implicitly critical of Government actions or policies' (Ozga and Lawn, 2014: 11). Instead of launching 'direct attacks' on the HMI, policymakers challenged HMI in terms of the lack of evidence used to form their judgements (Ozga and Lawn, 2014: 11) and argued for a 'tougher' role of HMI (Learmonth, 2000). HMI was increasingly seen as part of the problem with education by politicians, in particular by the Conservative Party (Exley and Ball, 2011; Lee and Fitz, 1997). These debates continued in the context of a 'growing disquiet about state schools from the late 1960s' (Elliot, 2012: 1), which became more visible in the late 1970s due to the critical voices of politicians, media, and

parents. The concerns were about educational standards linked to questions of competitiveness and national performance (Ozga, 1995: 27) as well as accountability and the effective use of resources (Hargreaves, 2010: 4). The autonomy of the local education authorities was questioned, and the 'professionalism and independence' of HMI were challenged in the context of the NPM (Ozga and Lawn, 2014: 11). For instance, according to the series of 'Black Papers' written by right-wing educationalists, politicians, and academics, published from 1969 to 1977, the rise of 'progressive education' in schools was claimed to be causing a decline in the standards of literacy and numeracy, as well as students' behaviour. Teachers were reported to be 'neglecting the basics and concentrating too much on informality' (Shaw, 2011: 8) and local authorities were reported as not sufficient to control them.

As an example of 'radically progressive' methods, at the William Tyndale Junior School in London in 1974-75, students were free to access all parts of the school including the staff room and using their 'open hours' as they liked. Newspaper reports referred to the school being 'out of control' with severe disciplinary problems (e.g., gambling, fire-starting) because students were 'too free' to do what they liked, including roaming the streets (Brighouse and Waters, 2022). At that time, the Inner London Education Authority endorsed such experimental schools, and allowed considerable autonomy to headteachers by replacing inspection with support for teachers and self-evaluation. The William Tyndale reports catalysed a loss of public confidence in the management of the system and local authorities' lack of control over education (Davis, 2002), leading to the 'Great Debate' after Prime Minister James Callaghan's speech at Ruskin College in 1976 (Brighouse and Waters, 2022). For Callaghan, not just teachers and parents but also government and industry should have an important part to play in formulating the aims of education. He also referred to other problems such as little attention to the basic skills of reading, writing, and arithmetic in the curriculum, and a lack of adequate skills for teachers to effectively discipline children. Callaghan concluded that the education system was out of touch with the fundamental needs of the country (Shaw, 2011: 9). According to Brighouse and Waters (2022: 13), this speech marked an end to the age of 'trust' and 'optimism' and opened up a new age of 'markets, centralisation, and managerialism', which has endured until today.

## 2.3 Education Reform Act 1988 and the four *pillars* of school evaluation

The 1988 Education Reform Act (ERA) enshrined into law a greater power of the central Government and sought to restrict the power of LEAs, as well as create more competition in

education by legitimising 'choice' for parents, in order to reduce 'inefficiencies' in the system and raise standards (Baxter, 2017: 22). As the power of local authorities was restricted, greater autonomy was placed with schools to manage their own budgets (Elliot, 2012). From the LEAs' point of view, it was not so much about giving more power to schools, but rather shifting power from LEAs to the Secretary of State for Education (Elliot, 2012). Just a few years later, in 1992, Ofsted was established, gradually replacing more independent HMI inspection with centralised, standardised inspections by Ofsted (Chitty, 2014). Central to these changes was the replacement of trust in the teaching profession (Brighouse and Waters, 2022: 13) with accountability and national inspection regimes that hold schools to account for delivering quality defined in terms of standardised test results and school league tables. Importantly for this thesis, the 1988 ERA introduced a new approach to evaluating schools that rested on the four *pillars* (see Figure 2.1), which are discussed in more detail below.
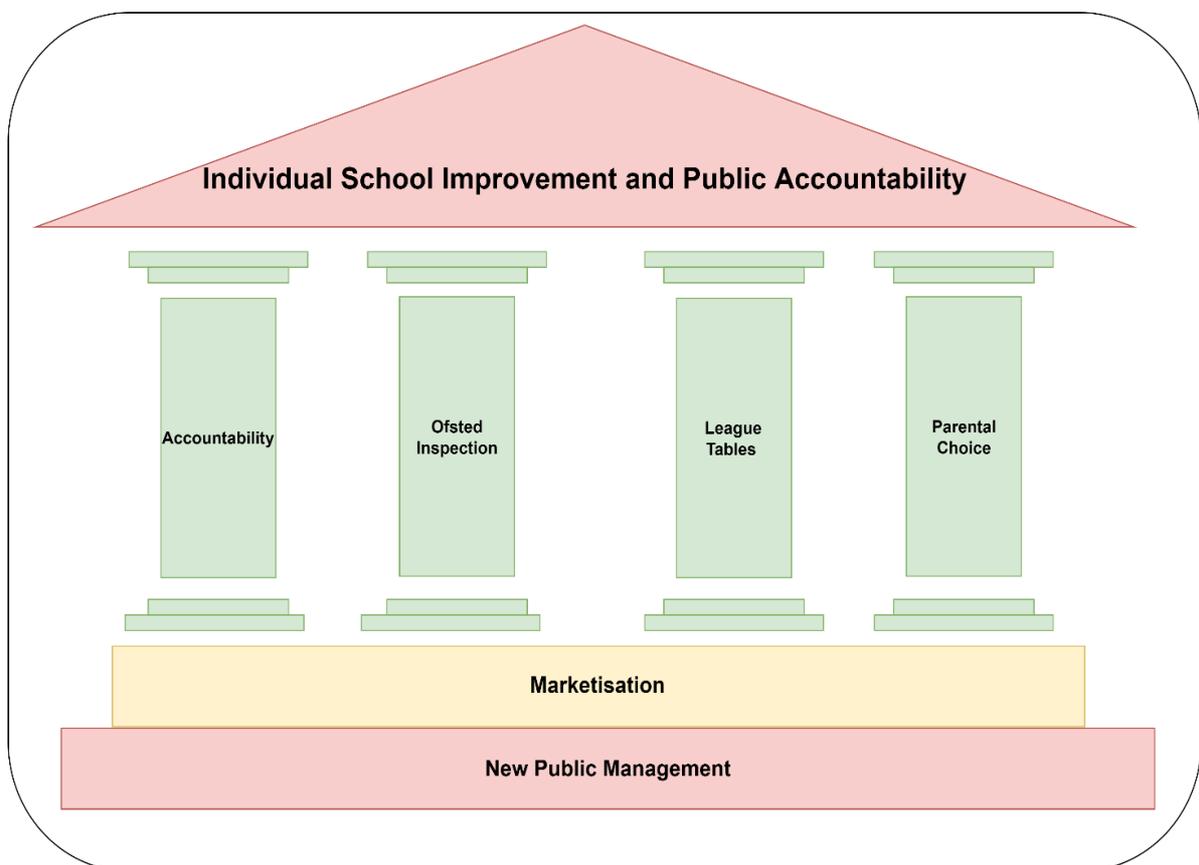


**Figure 2.1:** The *pillars* of the official approach to school evaluation

### 2.3.1 School Accountability

*Accountability* was encouraged since the 1980s (Brighouse and Waters, 2022) and intensified as a result of ERA 1988. With the establishment of Ofsted in 1992, accountability became a central, and centralised, objective of Ofsted inspections, within a market model in which parents are provided with information on school performance to inform their choice of a school (Exley and Ball, 2011; Gilbert, 2012: Wilkins, 2016). Whereas in the official approach, accountability is construed as a key driver of school improvement, it is also seen as a means of exerting pressure on schools and teachers (Day and Sammons, 2013). The negative effects of this pressure have included diminishing wellbeing and exodus of school staff. For instance, research by the NCSL (2006, cited in PwC, 2007) indicated that only one-third of retirements have been at normal retirement age or above in England. 43% of deputy heads and 70% of middle leaders did not aspire to take on the role as headteacher, for reasons that include 'accountability pressure'. This research also warned the authorities of potential future 'retention' and 'recruitment' crises of England (NCSL, 2006, cited in PwC, 2007). 73% of local authorities in England pointed out how their schools struggled to recruit suitably qualified staff (NfER, 2009).

Advocates of accountability, both within and beyond the UK, argue that accountability systems are needed to enhance students' performance (Schleicher, 2014). English advocates posit that accountability helps to ensure that school leaders are deploying resources effectively, and to assure the public that schools are doing well in their improvement (Day and Sammons, 2013; Gilbert, 2012). Altrichter and Kemethofer (2015: 32), through online survey data from 2,300 principals in seven European countries, found that headteachers who feel more 'accountability pressure' pay closer attention to the expectations communicated by inspectors and are more effective with regard to improvement activities. Similarly, Barber (2004) argued that, although accountability can cause pressure, it also helped deliver significantly improved outcomes since 1988. Advocates of accountability also suggest ways of making this pressure more manageable; for instance, by giving more autonomy to schools (Schleicher, 2012), making schools aware of the importance of accountability (Barber, 2004), and supporting them in better dealing with pressure (Elmore, 2006).

Some researchers focus on how accountability should be exercised. For instance, Elmore (2006) argued that there is an over-investment in testing and control and underinvestment in knowledge and skills needed to improve schools. This is because the success of an

accountability system depends on how it facilitates engaging the knowledge, skills and commitment of people who work in schools. However, performance-based accountability has been the dominant form of accountability in England since the beginning of the 1990s, with school league tables and outcomes-focused Ofsted inspection criteria. Interestingly, references to 'outcomes' were not emphasised in the Ofsted framework 2019 (Ofsted, 2019a), with the new focus of inspectors on the quality of the curriculum rather than schools' performance and position in the league tables.

Since accountability has been closely linked in England to Ofsted inspections which, in turn, have been preoccupied with student performance in high stakes tests as a key measure of school improvement, a circular argument seems to have emerged. Within this argument, accountability is construed both as a driver of school improvement (i.e., a *pillar* on which school improvement rests) and an overarching purpose of school evaluation. As discussed further below, problems arise when schools 'perform' for Ofsted inspectors (Perryman, 2009). Performing and fabricating evidence for inspectors means that schools withhold honest accounts of their successes or failures, which in turn defeats the purpose of school evaluation. As discussed in Chapter 3, development and social accountability to the public (students, parents, society, as well as the Government) should be the main purposes of school evaluation.

### 2.3.2 Ofsted Inspection

The Office for Standards in Education (Ofsted) was created by the Education (Schools) Act 1992, as an independent organisation, reporting directly to Parliament and responsible for inspecting all schools in England. Ofsted's main remit has been to check that schools are delivering better educational standards according to the externally set criteria and published quantitative performance information for every school (Courtney, 2012; Matthews and Ehren, 2017). Local authorities lost their inspection and advisory capacity over time. Instead, Ofsted inspected schools on a regular four-year cycle and made their report available to the public, with a letter sent to parents summarising the main findings (Elliot, 2012). This seems partly a response to the critics of HMI, as noted in 'HMI Today and Tomorrow' (1970: 24):

> The [HMI] report suffered from having to be written for a number of different readers whose needs were not the same. It often included information which some readers already knew (and had, perhaps, themselves supplied), while other readers had their appetite for judgments about the school's achievements whetted but not fully satisfied.

At first, Ofsted introduced a five-point grading scale in 1992 that grew to a seven-point school ranking system in 1996, which was later reduced to four: outstanding, good, requires

improvement, and, with drastic consequences, under special measures. The bottom category, synonymous with failure and inadequacy, often led to the departure of headteachers and other senior staff or governors in the school concerned. The burden on schools preparing for Ofsted inspections and the detrimental impact of inspections are widely evidenced in the literature (Chapman, 2000; Ehren et al., 2016; Perryman, 2009). For instance, just three years later after being established, Brimlecombe et al. (1995) collected questionnaire data from 821 teachers and interview data from 30 staff, who reported on the experience of school inspection as highly stressful. Similarly, Chitty (2014: 33) argued that the initial years of Ofsted inspections (1992 to 1997) were 'heavy' for schools. For Hofer, Holzberger and Reiss (2020: 14), between 1993 and 2000, Ofsted inspection was characterised by 'high accountability pressure, rather low support, low co-operation between the inspectorate and the school, and low adaptivity to the schools' strengths and weaknesses, and a focus on both output and process measures, negatively influenced GSCE scores.'

In response, in 2002, Ofsted published a document entitled *Reducing the Burden of Inspection*, promising a slimming down of the Ofsted requirement for schools to produce large amounts of evidence (MacBeath, 2005). In practice, however, the amount of documentation continued to grow, with the pace of reform accelerating after the victory of New Labour in 1997 (Chitty, 2014: 33; Thrupp, 2005: 42). Over time, Ofsted adopted a 'lighter touch' approach. For instance, the Education Act 2005 introduced short-notice inspections and school self-evaluation forms (SEF). The Education Act 2011 made outstanding primary and secondary schools exempt from routine inspections. Ofsted (2015) did not inspect those schools unless their performance deteriorates, and 'good' schools will not receive the intensive inspection, if not required. The key focus of Ofsted has turned to 'inadequate' schools and those 'requiring improvement' (Ofsted, 2015). However, it seems that these changes did not make a huge impact on the time spent preparing for the coming inspection (Ehren et al., 2016; Perryman, 2009). For instance, Ofsted's own survey in 2018 revealed that only 16% of teachers would 'carry on doing' their job as 'normal' if they knew Ofsted is coming. 79% of teachers noted Ofsted visits meant 'extra work', which is either 'unnecessary' (54%) or 'manageable' (25%). Ofsted's 2019 survey revealed similar results (Ofsted, 2019b). In November 2020, outstanding schools again became subject to routine inspection because many of them did not receive an inspection for a decade or longer (Roberts and Hill, 2019).

Another key issue regarding Ofsted concerns their impact on school improvement. Since its

inception in 1992, Ofsted adopted the strapline 'improvement through inspection' (Chapman, 2000: 57). However, the evidence as to how far it has succeeded in this aim appears inconclusive. On the one hand, in 2004, the first major evaluation of Ofsted's impact on the education system over 10 years since its inception was commissioned by Ofsted and carried out by the Institute of Education, University of London (Matthews and Sammons, 2004). The report pointed to 'considerable evidence' of Ofsted playing an important role as a 'catalyst' for improvement, particularly of weaker institutions. This was claimed to happen through the widespread improvement of 'the curriculum, assessment, leadership, and management of institutions and... the national strategies and standards for teaching' (p.154). The most and least effective schools made the greatest use of inspection findings, 'although for different reasons' (p.155). Because inspection provided the 'evaluation, leverage and accountability', it was argued that 'that have helped to embed such initiatives in educational practice' (Matthews and Sammons, 2004: 154). More recent research confirmed the benefits of Ofsted to improve schools (Altrichter, and Kemethofer, 2015; Ehren, 2016; Jones and Tymms, 2014).

On the other hand, some scholars found evidence of negative impact of Ofsted inspections on school improvement. For Ehren (2016: 2), 'despite this very favourable story about school inspections, there are also many caveats to be told'. Perryman (2009) found that inspections prompt schools to 'perform' during the inspection and fabricate evidence that inspectors wish to see. Rosenthal (2004) and Shaw et al. (2003) examined the effect of inspection visits on examination results, comparing the results a year after the inspection. Both studies found a small but significant worsening effect on exam results after inspection visits. Also, researchers interested in the perspective of heads and teachers on whether Ofsted is effective in 'securing school improvement', found that only 35% of schools felt that the benefits of inspection outweighed their negative effects. The rest of heads felt that inspection did not lead to improvement (Thomas, 1999). A recent annual report of HMI (2016/7) revealed that a significant proportion of schools graded as 'requires improvement' did not manage to improve following inspection: 33% of primary schools and 58% of secondary schools, previously graded as 'requires improvement', did not improve. Further, one in ten schools graded as 'requires improvement' were judged as 'inadequate' in the subsequent inspection. Ofsted's own surveys in consecutive years (2017, 2018, 2019b) concluded that just one in four teachers at most agree that Ofsted is able to be a 'force for improvement': 19% teachers agreed with this statement in 2017, 24% teachers in 2018 and 20% in 2019. What is more, only 17% teachers in 2017, 35% in 2018 and 18% in 2019 found Ofsted a 'reliable and trusted arbiter'

(Ofsted, 2019b).

A recent comprehensive, international systematic review of research evidence from the last 30 years (Hofer et al., 2020) sheds light on the reason why evidence is inconclusive and even contested in relation to the question of whether school inspection leads to improvement. Hofer et al. (2020: 17) argue that if we consider school improvement in terms of fulfilling the expectations of policies, then inspection does reach this aim. However, the literature does not provide a strong indication of inspection making a positive impact on 'school processes as well as input and context variables.' Therefore, they conclude that:

> it might be contested if inspection in its current form considerably contributes to sustainable school improvement that goes beyond an alignment of teaching practices and content with performance standards. (Hofer et al., 2020: 17)

Whether Ofsted guides improvement might be related to what we mean by 'school improvement'. Strategic behaviours (e.g., teaching to the test), as evidenced by Nelson and Ehren (2014), may look like improvement, but they can be at the expense of broader education. For instance, as evidenced by Greany and Higham (2018), 77% of primary and 83% of secondary school leaders agree that 'making sure my school does well in Ofsted inspections is one of my top priorities'.

Ofsted has been also criticised for applying 'superficial criteria' and ignoring the cultural, economic, or social factors that influence school performance (Richards, 2001, 2015), as well as employing methods which rely on 'pressure' rather than support for schools (Chapman, 2000). Although Ofsted's inspection approach is 'not likely to be the best system for engendering long-term improvement' (Perryman, 2009: 628); Ofsted remains in a powerful position and has many advocates (Gilbert, 2012).

### 2.3.2.1 Ofsted and standards for evaluating schools

As one of the four *pillars* in the official approach, Ofsted set national standards for evaluating schools. This section takes a closer look at Ofsted standards in order to consider their relationship to other available evaluation standards. No explicit 'standards' have been officially articulated by Ofsted, but an analysis of two recent Ofsted frameworks (2015, 2019a) reveals a number of criteria/evaluation questions guiding Ofsted inspections. These are listed in Table 2.1 below.

| Criterion/evaluation question | Ofsted (2015) | Ofsted (2019a) |
|---|---|---|
| Does the school comply with relevant legal duties, frameworks and national standards? | Inspectors will assess the extent to which the school or provider complies with relevant legal duties (2015: 3) and will evaluate in line with frameworks, national standards, or regulatory requirements (2015: 6, 17) | This evaluation question remained the same in the 2019 Framework |
| Relevant expertise of inspectors | Inspectors will inspect the type of provision for which they have the appropriate expertise and training to inspect the type of provision (2015: 4). | This criterion remained the same in the 2019 Framework |
| Professional standards of behaviour and positive working relationships | Inspectors and providers establish and maintain a positive working relationship based on courteous and professional behaviour (2015: 18). Inspectors must uphold the highest professional standards in their work, treating everyone they encounter during inspections fairly and with respect and sensitivity (2015: 16) and courtesy (2015: 17). Also, inspectors take all reasonable steps to prevent undue anxiety and minimise stress and respond appropriately to reasonable requests (2015: 17). | This criterion remained the same in the 2019 Framework |
| Public and government assurance as the purpose of evaluation | Ofsted provide assurance to the public and government that minimums standards of education, skills, and childcare are being met; that public money is being spent well; and that arrangements for safeguarding are effective (2015: 9) | This criterion remained the same in the 2019 Framework |
| Independent evaluation and objective judgement | Independent evaluation (2015: 6), evaluating objectively, being impartial and inspecting without fear or favour (2015: 17), and reporting honestly and clearly, ensuring that judgements are fair and reliable (2015: 17). | 'Fair' judgement has been replaced with 'valid' judgement achieved through the use of evidence, research, and inspector training |
| Diagnosis role of evaluators | Diagnosis of what should improve (2015: 6); providing challenge and the impetus to act where improvement is needed (2015: 10). | 'Diagnosis of what should improve' has been replaced by 'identify of what needs to improve' |
| Evidence-based evaluation | Inspectors will use all the available evidence to evaluate (2015: 23); a range of evidence (6); robust and clear evidence (2015: 17). | The phrase 'robust evidence' has been replaced by 'strong evidence'. |
| Providing information and 'informed choice' | The inspection provides important information to parents, carers, learners, and employers about the quality of education, training, and care being provided. These groups should be able to make informed choices based on the information published in inspection reports. Under the common inspection framework, readers will be able to compare different inspection reports quickly and easily, particularly where the reports are about provision for children or learners of similar ages (2015: 7). | This criterion remained the same in the 2019 Framework. |

| Coherence of judgements achieved through the use of consistent criteria | Having a framework (2015: 6, 7, 8, 22) and having consistent criteria (2015: 8) will lead to a coherent set of judgements (2015: 27). This will provide greater coherence across different providers that cater for similar age ranges and… will help inspectors to take comparable approaches to gathering evidence in different settings. Inspectors will comply with relevant guidance and codes of conduct, but they will always seek to be curious as well as compliant (2015: 8). | 'Researched criteria' was added in addition to 'consistent criteria'. |
|---|---|---|
| Communicate/inform judgement | Inspectors will maintain purposeful and productive dialogue with those being inspected and communicate judgements sensitively but clearly and frankly (2015: 17). | Purposeful and productive 'dialogue' has been replaced by purposeful and productive 'communication'. |
| Confidentiality of information (as far as possible) | Respect the confidentiality of information, particularly about individuals and their work (2015: 17) | This creation remained, with 'as far as possible' added to qualify the criterion of confidentiality. |
| Acting in the best interests and well-being of children | Inspectors will use all the available evidence to evaluate what it is like to be a child, learner, and other user in the provision. Inspectors will act in the best interests and well-being of service users, prioritising the safeguarding of children and learners at all times (2015: 17) and taking prompt and appropriate action on any safeguarding or health and safety issues (2015: 17).

Inspection is primarily about evaluating how well individual children and learners benefit from the education provided by the school or provider. Inspection tests the school's or provider's response to individual needs by observing how well it helps all children and learners to make progress and fulfill their potential. | 'Service user' has been replaced by 'learner' and the statement that 'all learners will receive a high-quality, ambitious education' has been added. |
| Improving schools through inspection | Improvement through setting standards, reporting on performance against relevant standards and raising expectations of performance. | Improvement through intelligent, responsible, and focused inspection and regulation. |
| Grading | Effectiveness of leadership and management Quality of teaching, learning and assessment Personal development, behaviour and welfare Outcomes for children and learners | 'Outcomes' was dropped, and the new grading criteria focused on: Quality of education Behaviour and attitudes Personal development Leadership and management |

**Table 2.1:** Criteria and evaluation questions guiding Ofsted Inspections

## 2.3.2.2 Ofsted standards in context: standards and principles developed by evaluation societies

In order to better understand Ofsted's 'modus operandi' presented above, it is important to note that a number of evaluation societies developed their own standards and principles to apply to all types of evaluation, irrespective of particular evaluation approaches. A set of well-known evaluation standards has been developed by the American Evaluation Association (AEA, 1994, 2018), which foregrounds 'systematic inquiry, competence, integrity, respect for people, and common good and equity' as guiding principles for evaluators. Another set of standards was developed by the Joint Committee of Standards for Educational Evaluation (JCSEE) for both personnel (1988) and programme evaluation (1994). The JCSEE's standards are educational evaluation-related standards. Although they were originally produced in collaboration between US and Canada, some European countries such as Germany, Austria, and Switzerland, either adopted those standards or adapted them to more diverse settings beyond educational settings. The JCSEE evaluation standards are widely cited by UK-based academics, for instance, Kushner and Norris (2007). They are conceptualised around 'utility, feasibility, property, accuracy' to guide a programme and personnel evaluation. The table below summarises some international examples of evaluation principles/ethics/standards/norms.

| The name of the organisation | Their work(s) |
| --- | --- |
| JCSEE | The personnel evaluation standards |
| | The program evaluation standards |
| AEA | Guiding principles for evaluators |
| UNESCO | The United Nations Evaluation Group: Norms and Standards for Evaluation |
| UKES | Guidelines for good practice in evaluation |
| | Evaluation capabilities framework |
| UK Department for International Development | UK Ethics principles for research and evaluation |

**Table 2.2:** Examples of evaluation principles/ethics/standards/norms

In the UK, the United Kingdom Evaluation Society (UKES) also published two sets of 'guidelines.' The first set consists of standards for 'good practice in evaluation' (UKES, 2013a), not only for 'evaluators' but also for 'commissioners', and 'participants' in an institutional self-evaluation. The second set comprises a framework for the conduct of quality evaluation, aimed at promoting a culture of professionalism and enhancing good practice in

evaluation (UKES, 2013b). The UKES standards are intended for use in any domain, discipline, or context. They are very detailed and comprehensive, and it is, therefore, puzzling why Ofsted did not appear to adopt them for school inspections. The next section offers a brief comparative analysis of Ofsted and UKES standards.

## Comparing Ofsted standards and UKES standards

A comparative analysis of Ofsted standards (2015, 2019a) and the standards developed by UKES (2013a, 2013b) reveals some notable similarities and differences. In terms of the content of the standards, both documents emphasise the expertise of evaluators, their commitment to evidence-based evaluation, and confidentiality of the information gathered. However, closer scrutiny of the language and explanations used reveals important differences. Regarding the expertise of evaluators, Ofsted refers to inspectors 'Having the appropriate expertise and training to inspect the type of provision' (2015: 4), 'upholding highest professional standards (2015: 7), and 'carry[ing] out their work with integrity, treating all those they meet with courtesy, respect, and sensitivity (2015: 8). There is no explanation of 'appropriate' expertise and training, and no specification of what constitutes 'highest professional standards' in this framework. The UKES (2013b) emphasises a broader and more precisely defined range of evaluator capabilities: 'evaluation knowledge, professional practice, qualities, and dispositions.' Furthermore, UKES (2013b) highlights the importance of qualities and dispositions (i.e., the personal characteristics that enable evaluators to function in difficult circumstances), evaluators' evaluation knowledge (i.e., knowledge base of philosophy and method evaluation brings from the social sciences, disciplines and professions and the specific knowledge) and evaluation practice (i.e., conducting a credible, valid evaluation, having the interpersonal and political skills to manage the process). By contrast, Ofsted (2015, 2019a) leaves the meaning of 'appropriate expertise' and 'professional standards' as undefined and abstract.

In terms of evidence-based evaluation, Ofsted's (2015) criteria refer to the collection and use of 'robust and clear evidence' (p.8), and inspectors' need to use 'all the available evidence to evaluate' (p.11). In the 2019 framework, Ofsted replaced 'robust evidence' with 'strong evidence'. However, in practice, the robustness and strength of evidence collected by Ofsted can be questioned, based on the length of inspections. Specifically, it is doubtful whether 'robust' or 'strong' evidence on the quality of education provided by the school can be collected in two days that a typical inspection takes (Ofsted, 2019a). The UKES' approach (2013a)

differs in their recommendation for a 'comprehensive and appropriate use of all the evidence' as well as for the conclusions of an evaluation that 'can be traced to this evidence' (p.3). This is important because it can increase the trust between the evaluator and the school's stakeholders.

For Ofsted (2015) on the other hand, trust does not seem to be of importance. Instead, the criteria refer to establishing and maintaining a 'positive working relationship based on courteous and professional behaviour' and conducting inspections in an 'open and honest way' (p.8). However, the openness and honesty encouraged by Ofsted standards are not likely to be achieved in an atmosphere where there is a focus on judgment rather than trust. The twelve-page document by Ofsted (2015) refers to judging and judgment 34 times, for example, a 'coherent' and 'common' set of judgments, 'consistent' criteria for reaching those judgments, written judgment, and 'making judgment'. Despite references to Ofsted 'evaluating without fear' and preventing 'undue anxiety' (p.8) the relentless focus on judgment creates an impression of the Ofsted judgment of a school as a 'fait accompli' that is highly likely to cause fear and anxiety. The key function here is a 'fair' use of the Ofsted inspection framework for comparability and checking schools' compliance with government policies and standards.

### 2.3.3 National curriculum, high stakes tests, and school league tables

The introduction of the National Curriculum was based on the claim that the curricula of some schools 'do not contribute as [they] should either to the efficiency of the schools or their responsiveness to national needs' (DfES, 1980: 1). The introduction of the National Curriculum facilitated standardisation through the Standard Assessment Tests (SATs) and General Certificate of Secondary Education (GCSE). With the introduction of school league tables, a lot was at stake for schools depending on SATs and GCSE results, leading to these examinations becoming 'high stakes' tests. The scores of individual schools were first reported nationally in a league table format in 1993 (Baxter, 2017). The publication of SATs results from primary schools soon followed. The logic behind school league tables is to make parents know which schools in their local areas are the best while they are exercising 'consumer' choice to send their children to schools that 'topped' the league tables and encourage schools to compete with each other to increase efficiency in the system (Baxter, 2017). Over time, school improvement came to be defined mainly in terms of children's exam results and school rankings.

School league tables are, however, much-contested as evaluative mechanisms because of the

negative impacts on students, schools, and even parents. Concentrating on test results and league tables divides already segregated communities, with parents competing for the best schools for their children (Allen and Vignoles, 2007; Ball, 2003; Gewirtz et al., 1994). Researchers have also found data on the negative impact of high stakes tests and league tables on student learning. According to Torrance (2011), governments tend to over-concentrate on high stakes test results, but higher test scores do not always mean better learning. On the contrary, test scores may mask falling standards, because of restricted curriculum and reduced quality of learning due to 'teaching to the test'. The 2019 Ofsted Framework aimed to solve this 'narrowed curriculum problem' by focusing on the quality of the curriculum rather than test scores (Ofsted, 2019a). The competition principle underlying school league tables has also been applied internationally, in the form of the 'Programme of International Student Assessment' (PISA), developed by the Organisation for Economic Co-operation and Development (OECD). Alexander (2012) notes that with the growing international trend to measure the effectiveness of education systems through comparative studies such as PISA, some countries 'borrow' policies or even the entire national curricula of successful countries without thinking about their own culture, values, and needs.

### 2.3.4 Parental choice in the 'school market'

The fourth *pillar* in the official approach has been the marketisation of schools and the emphasis on parents' choice of schools for their children. While the 1980 Education Act introduced 'parental preference' of school and parents were encouraged to serve on governing bodies, their 'power' was increased in the 1984 Green Paper, *Parental Influence at School* (gov.uk, 1984)*,* which highlighted the vital role and responsibilities of parents to play in the education system. Then, with the 1988 ERA, marketisation became more visible, legitimising open enrolment ('parental choice') and formula funding ('pupil-led funding') through the local management of schools. Kenneth Clarke, the Education Secretary between 1990 and 1993, declared that parents were a key 'driver' of the new system (Wilcox and Gray, 1995). With open enrolment, parents were given the right to choose the school they wanted their child to go to and could appeal if they were not accepted by the school they chose. They started to be informed by Ofsted inspection reports and school grades as well as school performance ('league') tables from the early 1990s. With pupil led-funding, schools were funded for the number of pupils on roll. The intention was that parental choice and per capita funding would induce competition between schools, raising educational outcomes (Glennerster 1991, cited in Burgess et al., 2007). According to Hammersley (2007), seeing parents as 'customers' and

giving them choice stems from the NPM, and government attempts to set up quasi-markets in education to maximise efficiency.

Open enrolment, as one of the key features of marketisation, has been widely critiqued because it increases inequality (Hatcher, 2011; Hicks, 2015; Thrupp, 2005) and best serves middle class families. For instance, Gewirtz et al. (1994: 9) described 'middle class parents' as 'skilled choosers' (unlike 'disconnected choosers', i.e., working class parents) and demonstrated how choice is exercised by middle-class parents because they can easily access resources and networks that enable choice. Ethnic segregation has seen another negative consequence of parental choice (Burgess et al., 2005). Another problem with open enrolment is related to schools' admissions criteria and the availability of places. The research of Burgess et al. (2007: 288) examined the data set of over a million students in England and found out that the allocation of children to particular schools is governed by 'the interaction of demand- and supply-side of selection and the feasibility of choice'.

In a similar vein, pupil-led funding as another key feature of marketisation was not without critique. For instance, a three-year study by Gewirtz et al. (1994: 5) concluded that 'resources are flowing from those children with the greatest need to those with least need' because of the processes of de-comprehensivisation triggered by parental choice. They noted that 'growing inequality of access to the quality of provision necessary for children to succeed educationally' should not be surprising to 'the architects of the UK market' (p.5) because they are not committed to 'needs-based equity'. Further, Gewirtz et al. (1994: 14) pointed out that 'success in the marketplace is not primarily a function of family motivation but rather of parental skill, the perceived raw-score potential of the child, and, to some extent, pure chance'.

Marketised forms of education in state education were also expanded through establishing different types of schools in the name of the 'diversity' of schools to choose from and those schools were given 'greater legal and financial freedoms to govern themselves' (Wilkins, 2017: 172-177). For instance, between 1986 and 1990, City Technology Colleges were created. They were directly run by the Department for Education (DfE) and funded by private sponsors. The 1988 Education Act also encouraged secondary schools to 'opt out' of LEA control by having grant-maintained status. The 1933 Act extended choice and diversity through the 'financial encouragement of specialist schools and through the acceleration of opt out' (Ozga, 1995: 30). More recently, with the 2010 White Paper, the marketisation of education was accelerated through the academies and free schools programme, both reviving the grant-maintained model,

which not only invites private sponsors to sponsor and run (through governing bodies) academies and free schools, but also encourages the formation of federations of schools and Multi Academy Trusts (Hicks, 2015). Whilst academy schools are state-funded schools which are directly funded by the Department for Education and independent of local authority control, Multi Academy Trusts (MATs) operate more than one academy. MATs are legal entities and are run like a company by a board of directors, responsible for the management of multiple schools under a trust that administers land and building ownership, sets the curriculum and admissions policy, manages budgets, and employs staff (West and Wolf, 2018). Whereas Ofsted inspects academies and local authority-controlled schools, they cannot inspect MATs. Because MATs and standalone ('converter') academies receive their funding directly from the central government rather than local authority, the outcome is an increasing centralisation of schools in the name of decentralisation (Brighouse and Waters, 2022). This goes hand in hand with 'centralised accountability' (Wilkins, 2017: 178). According to Wilkins, MATs subscribe to 'market-oriented approaches to school governance' (p.179) in which the school's autonomy is undermined. Their formation is driven by the political aim of managing 'individuals and institutions that are otherwise beyond the scope of ordinary governments' (p.178). As Wilkins explains:

> Arguably local government is no less technocratic or bureaucratic than the MAT model, to the extent both make themselves accountable to others through the setting of performance indicators, the establishment of benchmarks, the administration of audits, the delegation of management overheads and the contracting out of services. The key difference here is that local government is vulnerable to capture from political opposition and deliberation as a condition of its democratic, open-ended organisation. In other words, the power to govern under democratic rule is not reserved for experts, professionals, and technocrats exclusively, making local government 'politics' unwieldly and unpredictable. Local government politics is risky in so far as it cannot be comfortably squared with the kind of rational-consensus, techno-managerialist approach to school governance favoured by politicians and policymakers who counterpose 'quality' and 'democratic accountability'. (Wilkins, 2017: 179)

In the name of 'quality' rather than 'democratic accountability', the 2016 Education and Adoption Act made it mandatory for maintained schools (under LA control) which received the 'inadequate' grade from Ofsted to join a MAT. For Wilkins (2016), this government-led transformation towards 'academisation' undermines democratic accountability as it removes 'the steering capacity' of the local government to 'hold schools to account on behalf of the communities they serve'. As a result, school governors are expected to take on responsibility for 'democratic accountability' through 'strategic oversight' on schools. However, democratic accountability cannot be guaranteed by governors because they are constrained by the 'calculus

of the market' and, in addition, are 'overloaded with prescriptive accountabilities that disproportionately serve the interests of the funders and the regulatory body' (Wilkins, 2016: 1-2).

The education system in England is thus becoming more like a marketplace, with parents, in theory, being able to choose the best schools for their children. However, according to Hicks (2015), the idea that the choice will drive school improvement is extreme. For Ball (2003: 219), marketisation and competition is a form of 'entrepreneurial control'. For other researchers (Hatcher, 2011; Hicks, 2015; Thrupp, 2005), the competition logic behind marketisation has a negative effect on school's self-evaluation, as it leads to school self-promotion rather than realistic school self-evaluation.

## 2.4 School self-evaluation

Between 1960 and 2000, self-evaluation has been 'revisited, embraced, and then forgotten as new ideas and more pressing priorities emerged' (MacBeath, 2005: 10). Self-evaluation can be traced back to 'GRIDS', a form of self-evaluation using a set of criteria for schools to review their own performance in the 1970s (MacBeath, 2005: 10). Teachers in UK schools during the 1970s and 1980s often assumed that they alone were responsible for the quality of their practice and therefore had the power to change it if they wished without any structural constraints (Elliot, 1993; Shaw, 2011). The use of GRIDS was gradually phased out as teachers were better at identifying improvement rather than bringing about improvement (Fidler et al., 1997: 63) and as the new inspection regime raised the profile and stakes of external evaluation (MacBeath, 2005).

In 1997, the New Labour government put self-evaluation back at the heart of school improvement (MacBeath, 2005) and Ofsted endorsed self-evaluation through *School Evaluation Matters* in 1998 (Ofsted, 1998). Therefore, from 1999 onwards the demands on schools increased towards self-evaluation, with a range of self-evaluation forms such as 'PICSIs' (pre-inspection information) and 'PANDA' (performance and assessment data). Each indicator needed to be accompanied by evidence, summaries of strengths, areas for improvement, and priorities for future development. The proliferation of indicators and evidence that schools had to prepare, typical of NPM approaches, was criticised for being a burden and an imposition, despite support for the idea of self-evaluation (MacBeath, 2005).

Although successive governments in England supported school self-evaluation they also

promoted, directly or indirectly, evaluation processes that measure and assess outcomes of importance to the government and Ofsted (Ehren, 2016; MacBeath, 2005; Ritchie, 2007). For instance, in 1999, Ofsted highlighted that self-evaluation is an important part of the inspection. This was further endorsed through the Framework for Inspecting Schools (Ofsted, 2003) and the Department for Education and Skills recommendation for self-evaluation to be put at 'the heart of the inspection', as 'the most crucial piece of evidence available to the inspection team' (DfES, 2004: 24). However, making school self-evaluation inextricably linked to the inspection regime made it 'more closely associated with accountability than improvement' (Ritchie, 2007: 85).

The Education Act 2005 introduced a formal requirement of self-evaluation to schools based on the guidelines developed by Ofsted (Chitty, 2014). The 'Self-Evaluation Form' (SEF) consisted of 35 pages which schools were required to complete as a 'summative document, intended to record the outcomes of an ongoing process of rigorous self-evaluation' (Ofsted, 2005: 3). SEF covered a set of questions about characteristics of the school; views of learners, parent/carers, and other stakeholders; achievement and standards; personal development and wellbeing (of learners); the quality of provision; leadership and management; overall effectiveness and efficiency. The school was also required to grade aspects of work on a four-point Ofsted scale (outstanding; good; satisfactory; inadequate). It seems that an imposed, over-prescribed SEF template went against the very principle of self-evaluation, which is predicated on providing opportunities for schools to freely discuss their own perceived strengths and weaknesses and develop their own templates and development plans (MacBeath, 2006).

Research on the usefulness of SEF produced mixed findings. For example, a study by Bubb and Earley (2008), which focused on school improvement plans of 38 schools in England, revealed that schools planned their improvement by filling out the SEF form before the inspection. Some schools were left feeling overwhelmed by all that they had to improve. Overall, however, Bubb and Earley (2008: 5) concluded that 'completing the SEF can make people realise the need to change the strategic direction of the school and rethink their improvement priorities, whilst the format of the SEF can reinforce the logic of the link between improvement plans and those for staff development and training.' Other researchers put forth strong criticisms of SEF. For instance, MacBeath (2006: 5) argued that the SEF is 'not self-evaluation'. Brady (2016) pointed out that, because SEF was based on Ofsted criteria, self-evaluation became a form-filling exercise, corrupting the very notion of *self* evaluation, and

was not a genuine attempt at improving practice.

Currently, schools in England still undertake self-evaluation (Matthews and Ehren, 2017). However, with the Coalition Government coming to power in 2010, the form for self-evaluation (SEF) provided by Ofsted was scrapped (Chapman and Sammons, 2013) and schools are allowed to use any method, models, and criteria for their evaluation as optional (Courtney, 2012). The scrapping of the obligation for schools to complete the SEF received a mixed response: some school leaders structure their whole school self-evaluation process around SEF; some use it because Ofsted expected them to do so; some school leaders have never used the SEF and relied on their own, internal self-evaluation processes (Chapman and Sammons, 2013). Although schools are now free to use their own self-evaluation formats (Matthews and Ehren, 2017), SEF was reported to be used by 61% of respondents in a study by Courtney (2012) who agreed that they will retain the level of emphasis they place on SEF and the Inspection Framework, as well as 28% respondents who claimed they would increase the focus on it. Similarly, Matthews and Ehren (2017) pointed out that despite the advocacy of more bottom-up approaches, many schools still choose to base their self-evaluation on the aspects covered by Ofsted inspectors.

Although self-evaluation in England is a demanding procedure, a review of international research published between 2007 and 2015 suggests that self-evaluation can lead to sustainable school improvement and increase student achievement (Nelson et al., 2015). Similarly, the OECD (2015: 130) confirms the importance of self-evaluation in promoting school improvement. Studies by Bubb and Earley (2008) and Ehren (2016) also highlight that careful self-evaluation has direct effects on continuous school improvement.

Whilst conducting self-evaluation can be useful and worthwhile, Brady (2016: 523) notes that it does not 'live up to its name' if the criteria used for self-evaluation are not internally generated but externally imposed. Similarly, Ehren (2016) highlights that schools often implement self-evaluation in response to school inspections. In a similar vein, Janssens and Amelsvoort (2008) claim that, either mandatorily or voluntarily, schools tend to use inspectors' criteria, leading to a situation where self-evaluation becomes standardised and aligned to the demands of external evaluation. There are also schools that resist engaging in self-evaluation. For example, Hall and Noyes (2007) found that internal evaluation was seen in some schools as a bureaucratic exercise; school leaders were cynical about the process; teachers were complaining about the workload and change fatigue. Also, Bubb and Earley (2008: 12)

emphasise that school self-evaluation should be an integral part of improvement focused on priorities and staff development rather than a 'game that people play in order to 'pass' their inspection'. Interestingly, Ofsted's (2019a) framework notes that inspectors will not look at internal data. Instead, they will ask leaders at various levels, what they understand about progress and attainment in the school or subject. They will gather their own data by focusing on what is being taught and learned through lesson visits, work scrutiny, and conversations with learners. This new approach might lead to two consequences. The first is that fewer schools would be interested in self-evaluation as it is no longer required by Ofsted inspectors. The second is that this new approach would allow schools to be more open to seeing their own weaknesses because it would be used internally, rather than viewed inspectors. Because of the pandemic, there are no empirical studies yet on the impact of this new approach on the conduct of school's self-evaluation.

## 2.5 Methodological underpinnings of the official approach to school evaluation

Any type of evaluation is guided by a methodology and evaluation techniques used to acquire knowledge about the quality of a programme or service provided by a school (Christie and Alkin, 2013). In other words, approaches to school evaluation have methodological underpinnings, for example, based on quantitative or qualitative approaches. Using scientific research methods and techniques may improve the rigour in evaluations (Christie and Alkin, 2013). Because research methodologies are debated and contested, different approaches to school evaluation are also subject to debate. The evaluation may also be influenced by values and political considerations, such as those promoted by the NPM (Norris and Kushner, 2007) discussed above. For example, as a result of NPM, school performance statistics and quantitative data reported in schools' league tables became the main indicator of 'top' and 'underperforming' schools.

This section discusses the methodological underpinnings of the official approach to school evaluation, which tend to rely more on quantitative than qualitative data. Broadly speaking, quantitative approaches analyse whether a programme or practice is effective in bringing about the desired effects and to explain and predict outcomes, using experimental methods and positivist assumptions of an 'objective' reality that can be 'discovered' by the evaluator (Cohen et al., 2007). Qualitative approaches seek to improve the understanding and meaning of practice. While quantitative approaches to school evaluation may use student achievement data

in high stakes tests, or surveys to provide findings on the correlation between teaching and school outcomes, qualitative approaches follow a more exploratory approach, for example by interviewing school staff, parents, and students about specific problems or good practices. Since the rise of 'evidence-based' policy and practice in the 1990s (Hammersley, 2007; Hargreaves, 2007), medical models have also been promoted based on Randomised Controlled Trials (RCTs), discussed later in this section.

Looking at the official approach to school evaluation discussed above, Ofsted inspectors collect 'evidence' on examination and test results, analysis of school documents, questionnaires for parents, interviews with school staff, and observations of teaching. As stated by Thrupp, (2005: 16), an Ofsted inspection could be considered as a 'case study', although it is methodologically 'flawed', because of a narrow sample and problematic validity of the 'evidence' collected by inspectors, especially where schools seek to impress the inspectors. In fact, Fitz-Gibbon and Stephenson (1996: 17) argue that Ofsted's methods have 'failed to meet even the most elementary standards with regard to sampling, reliability, and validity'. Thrupp (2005) argues that official case studies of improving schools (Ofsted inspections) are usually too 'tidy' to be 'true' and fail to provide a sense of the day-to-day struggles and tensions conveyed by independent accounts of schools.

Most official school evaluation is based on school performance statistics in the form of quantitative data on pupil performance in high stakes tests. This approach is reductive in not taking into account the role of culture, context, socio-economic, and other complex factors that influence test results (Hammersley, 2007). Since the famous debate started by the TTA lecture in 1997 delivered by David Hargreaves, medical research models have been promoted as part of 'evidence-based' policy and practice within a 'what works' approach (Biesta, 2007). Post-2010, RCTs became a favoured official source of evidence-based research; and this method is used to generate knowledge of specific interventions tested to find out whether they improve pupil outcomes. For example, Goldacre (2013) points out that RCTs are good at showing the effectiveness of specific interventions, even though RCTs may be criticised for not being able to understand why an intervention has worked and in which context. The underlying assumptions of evidence-based policy and practice are positivist (Berliner, 2002; Elliot, 2001, 2007). Elliot (2007) sees this as a limitation for educational research and thus recommends robust case study research designs as the most appropriate mode of educational research for informing practice.

The debates on methodologies that can be utilised to evaluate schools are not limited to England and have included broader discussions of the nature of social science. Here, Robert Stake's (1986) discussion of a US project called *The Cities-in-Schools* programme is important. The evaluators of this project concluded that the programme did not deliver on many of its promises and did not make an impact worthy of the US government investment. Stake argues, however, that this conclusion was reached because the evaluators were committed to 'social science as quantification' and consequently the programme evaluation sought to find relationships among quantitative indicators of program operation, rather than theory-disciplined knowledge base (Stake, 1986: 133). Stake (1986: 133) argues that social science methods can be used to understand a complex 'array' of what a programme has accomplished, but if used just to find statistically significant differences, the results will be simplistic. For Stake, the way forward is not to decide which approach is more appropriate, but to be aware of the utility of 'generalistic' (quantitative, positivistic) versus 'particularist' (case study) evaluation designs. In conclusion, there is not one appropriate method to evaluate a school or a programme and a combination of different sources of evidence can support effective and useful school evaluation.

## 2.6 Alternative approaches to school evaluation

Although there are many critics of the official approach to school evaluation and its underpinning methodologies, few alternatives are presented in the literature. As explained by Chapman (2000: 57), alternative approaches have been limited because of research which demonstrated that alternatives such as further professional development of teachers led to 'negligible' school improvement (Fullan, 1991; Hopkins, Ainscow and West, 1994). The alternative approaches could be divided into 'moderate alternatives' (suggesting improving the current approaches) and 'extreme alternatives' (suggesting more radical ways such as scrapping the current evaluation systems or even avoiding school evaluation altogether) (see Table 2.3).

|  | **How to evaluate?** |
|---|---|
| Moderate Alternatives | Context-Specific Model |
|  | School to school peer review |
|  | Focus on school self-evaluation and capacity building |
|  | Sharper usage of evidence |
| Extreme Alternatives | Scrapping of external evaluation and committing a professional accountability system |
|  | Romantic approach |
|  | Goal free Evaluation |

**Table 2.3:** Alternative approaches to school evaluation

The *Context-Specific Model* focuses on the evaluation of activities such as assessment, target setting, inspection, performance management, staffing, and funding on social and organisational complexity (Thrupp, 2005: 112). Researchers such as Davis and Martin (2008), Harris and Chapman (2004), and Townsend and Avalos (2007) suggest this model as a way of moving away from a 'one size cannot fit all' evaluation.

*School to school peer review* has been developed to support the self-improving school system in England since 2010, within a decentralised approach to school evaluation (Matthews and Ehren, 2017). Peers are neighbouring colleagues, and peer review teams involve two or three schools (Matthews and Headon, 2015) who 'come together and spend time in each other's contexts to review practice, share expertise, recommend strategies for development and challenge each other to achieve continuous improvement' (Peer Challenge, nd). Peer review can reduce the frequency and intensity of inspection (Challenge Partner, 2015; Earley, 2013; Matthews and Headon, 2015), as well as allowing schools to focus more on improvement than accountability because it is more about supporting and challenging schools. Peer review can also complement self-evaluation through validation and calibration of the findings (Matthews and Headon, 2015; Ozsezer, 2016) as it increases self-knowledge in participating schools. This approach relies on peer review which, according to Matthews and Ehren (2017: 47), allows schools to take a 'greater ownership of their quality assurance, not only through self-evaluation but by exposing their work to the perceptions of trusted peers'. Matthews and Headon (2015) suggest that peer review deserves to be recognised and encouraged explicitly through national policy.

*Focusing on school self-evaluation and capacity building.* A recent OECD (2015) report explains that schools are well placed to analyse their own contexts, including performance and areas for improvement. Internal evaluations can lead to sustainable school improvement (Nelson et al., 2015) by building self-evaluation and reflective capacity throughout the school (Blok et al., 2008; McNamara and O'Hara, 2009).

*Sharper usage of evidence.* According to Levitt et al. (2010: 31), 'some degree of audit, inspection and scrutiny will undoubtedly continue, as it has become part of wider democratic governance, especially where matters are technically or bureaucratically complex.' However, the use of more precise evidence can lead to greater improvement to practice. Therefore, Levitt et al. (2010: 31) suggest that policy and practice should be informed by a better understanding of how audit, inspection, and scrutiny can use evidence more effectively to form judgements

and promote improvements in the quality of public services.

*Goal Free Evaluation* leaves to evaluators the responsibility for determining which programme outcomes to examine and rejects the objectives of the programme as a starting point (Scriven, 1997). This allows the evaluator to identify the real accomplishments of the programme using a more qualitative approach to describe events, reactions, and interactions. Goal Free Evaluation also allows the evaluator to adapt to stakeholder needs and concerns.

The *'romantic approach'*. Whereas the existing elements of the official approach to school evaluation (Ofsted inspection, league tables, target setting, performance indicators) remain problematic, Thrupp (2005: 49) notes that removing some of them is not being discussed. Even if these elements are criticised, evaluating schools is necessary. For instance, Chapman and Earley (2010) argue that 'there should always be a role for the external eye or outsider perspective' (p.724). Removing any of these elements could be called a 'romantic approach' that is not realistic. Apple and Beane (1999: xi-xii, cited in Thrupp, 2005: 7) argue as follows, based on the US context:

> Think of it: no 'league tables'; no pre-specified national curriculum or national testing programme, somewhat fewer worries about the image in the face of a competitive school market. Yet this totally romanticises the situation that [US] educators face…

*Committing to a professional accountability system.* Abolishing the external inspection system (for teaching staff) and focusing instead on professional accountability was introduced in Finland in 1991 (Penninckx and Vanhoof, 2017). In this model, the main focus of monitoring is not the task of the bureaucracy but the professional community, i.e., the teaching staff, who are supposed to be the best judges of how to ensure quality education. The model relies on the expertise and ethical code of the teaching profession. Its dominant approaches to monitoring are self-evaluation and peer review (Ehren, 2016; Penninckx and Vanhoof, 2017).

## 2.7 Conclusion

Although the 1988 ERA stated that one of its objectives was to give more control to schools over their own management; the freedom of teachers and schools was in fact greatly reduced within the new approach to evaluating schools. Schools gained control over their own budget but could not control their curriculum, in response to their local context and specific needs of their students. The four *pillars* of the official approach to school evaluation seem to serve mostly the needs of the Government and external stakeholders who wish to be assured that schools are under control and meet the centrally defined standards as a way of justifying their

receipt of public funds. Whether Ofsted inspections can or have improved schools is not clear, what is, however, clear is that improvement has been narrowly defined in terms of student performance in high stakes tests, which became the main 'metric' in school evaluation. As Lee and Fitz's (1998, cited in Ritchie, 2007: 88) emphasise, whilst external inspection may be an appropriate instrument for judging schools if this is what intended, it cannot, on its own, lead to improvement. According to the critiques of the official approach discussed in this chapter, the ideas of marketisation and accountability (which seems to have become an end in itself rather than a means to school improvement or an opportunity for schools to offer honest accounts of their provision), together with the NPM's focus on targets and control, may not be the best foundation on which to build an approach to the evaluation of schools that is fit for purpose. The following chapter explores developments in the field of evaluation to identify alternative purposes of school evaluation and their respective alternative *pillars* and alternative *foundational ideas* (Christie and Alkin, 2013).

# CHAPTER 3: CONCEPTUAL FRAMEWORK

## 3.1 Introduction

As noted in Chapter 1, ever since evaluation was formally established, approaches to evaluation have been influenced by the 'social and political climates of the day', with its funding and the 'sets of questions it studies often changed with political tides' (Shadish and Luellen, 2005: 186). The rise of the official approach to school evaluation in England has relied on the four *pillars* (see Chapter 2) associated with the 'political climate' that also gave rise to New Public Management (NPM) and the 'audit society' (see Power, 1997). Since this approach can be seen as a product of a particular social and political 'climate', its purpose, design and methodology, as well as its consequences, need to be scrutinised. Whilst Chapter 2 focused on the four *pillars* introduced by the Education Reform Act (ERA) 1988 and core debates on its impact on schools, this chapter seeks to develop a conceptual framework for rethinking school evaluation in England by exploring theoretical debates in the field of evaluation.

The chapter starts with Christie and Alkin's (2013) 'Evaluation Theory Tree' as a helpful metaphor that captures the development of the many branches of evaluation, as well as their roots in the *foundational ideas* of *social accountability*, *social inquiry* and *epistemology* (Section 3.2). Drawing on the key distinctions in the field of evaluation, Section 3.3 then lays out the similarities and differences between 'traditional' and 'developmental' approaches to evaluation, including their distinctive methodologies, points of difference in relation to the use of evaluation findings, as well as contrasting perspectives on valuing (judging) and values. Section 3.4 discusses democratic evaluation (MacDonald, 1974, 1976; Norris, 2015; Picciotto, 2015; Simons, 1987), a distinctive approach developed in the UK. The discussion then moves to the conceptual framework for rethinking school evaluation in England (Section 3.5). The conceptual framework draws on concepts from developmental evaluation and democratic evaluation that are helpful in rethinking the purpose, method, use and valuing in school evaluation, leading to a recasting of the *pillars* of the official approach (Figure 2.1 and Figure 3.2 below).

## 3.2 Evaluation Theory Tree (Christie and Alkin, 2013)

To systematically capture the development of the many branches in the field of evaluation, Christie and Alkin (2013) developed an 'Evaluation Theory Tree' (see Figure 3.1 below). As Christie and Alkin explain, in response to doubts about the effectiveness of government-funded

social programmes in the USA, since the 1970s various evaluation theorists 'prescribed' specific evaluation methodologies. The Evaluation Theory Tree is a result of a review of over 30 different evaluation approaches (and theories), categorised according to their *foundational ideas* (roots, foundations), with three 'branches' grown from these *foundational ideas*. Christie and Alkin included in their Evaluation Theory Tree the key theorists whose influence on the field of evaluation has been lasting. Importantly, the branches are not independent of each other, and some theorists could be included in more than one branch. Christie and Alkin's grouping of particular theorists on a specific branch reflect the primary emphasis of these theorists. Grouping theorists in branches has been intended to 'reflect a relational quality between them' (Christie and Alkin, 2013: 13). Whilst Christie and Alkin foreground American theorists, this chapter will also discuss, where appropriate, theorists who have contributed to advances in the field of education in the UK.
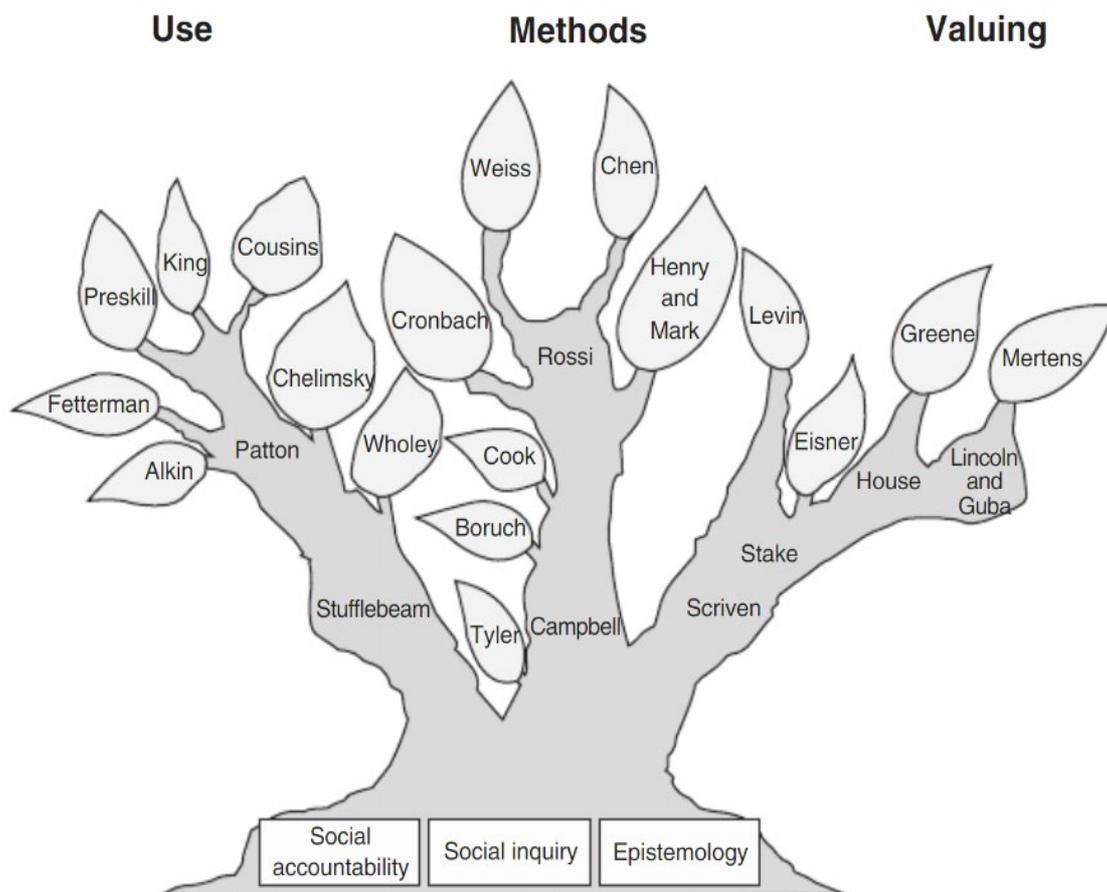


**Figure 3.1:** Evaluation Theory Tree (Christie and Alkin, 2013: 12)

As can be seen from the above figure, the primary emphasis of evaluation theorists such as Patton and Stufflebeam is on the use of findings from evaluation. Theorists such as Campbell, Tyler and Cronbach focused primarily on methods of evaluation, whilst Scriven, House and Stake were interested in the issue of valuing (judging) associated with evaluation (Christie and Alkin, 2013). It is also important to note that, for Christie and Alkin (2013), *foundational ideas* for evaluation stem from *social accountability*, *social inquiry* and *epistemology* and can, therefore, provide important conceptual tools that can serve two purposes in relation to school evaluation in England. Firstly, they can provide a 'counter-balance' to the official approach, which reflects the 'political climate of the day' (Shadish and Luellen, 2005: 186). Secondly, they can help us to rethink school evaluation by refocusing it on broader questions of *social accountability*, *social inquiry* and *epistemology*.

It is also important to note that the figure above does not feature theorists based in the UK such as Barry MacDonald, Helen Simons, Nigel Norris, Ray Pawson and others. This figure is the fourth, updated version of the 'Evaluation Theory Tree' by Christie and Alkin (2013: 12). Although this framework is useful in capturing the divergence and similarities of the theories, it is limited to North American theorists. For example, Christie and Alkin removed Barry MacDonald and John Owen from the 2013 version of the diagram 'because both reside outside North America (Great Britain and Australia, respectively) and their writings relate to work in these countries' (2013: 51). Consequently, Christie and Alkin's Evaluation Theory Tree needs to be approached with caution, as representing what they considered to be 'pioneering figures' in the field of evaluation and, as such, it is not complete or exhaustive. Since this thesis focuses on approaches to school evaluation in England, it includes important 'pioneering figures' in the field of democratic evaluation such as MacDonald (1974, 1976) and scholars who have developed his ideas (Norris, 2015; Picciotto, 2015; Simons, 1987). As can be seen in Chapter 9, democratic values have been proposed as a *foundational idea* by some of the experts who participated in this study.

The diagram also lacks theorists from low- and middle-income countries and the so-called 'developing countries'; even though Carden and Alkin (2012) attempted to address this issue. Their exploration led them to conclude that many existing practices in these countries are yet to be formalised into theory. In conceptualising 'developmental' evaluation, Patton (2011) was influenced by his work with indigenous communities. Patton's developmental evaluation is distinctive because it 'situates the evaluator inside a program' (Carden and Alkin, 2012: 110).

Carden and Alkin (2012) also note that the so-called 'indigenous' evaluation methodologies often 'express a gap in addressing local values in the methodologies developed in industrial countries that most closely reflect them' (p.111). They are, therefore, 'region specific' and constructed with the predominant concerns of use or valuing. A more inclusive approach to the Evaluation Theory Tree is presented in Table 3.1 below.

| Use | Methods | Valuing |
|---|---|---|
| Developmental Evaluation<br>*Patton*<br>(Adapted) | | Rapid Rural Appraisal,<br>Participatory Rural Appraisal<br>*Chambers*<br>(Adapted) |
| Outcome Mapping<br>*Earl, Carden & Smutylo*<br>(Adapted) | | Systematization<br>*Selener, Tapella et al.*<br>(Indigenous) |
| Most Significant Change<br>*Davies and Dart*<br>(Adapted) | 3ie: International Initiative on<br>Impact Evaluation<br>*Center for Global Development*<br>(Adopted) | |
| Citizen Report Card<br>*Public Affairs Centre, India*<br>(Indigenous) | | |
| African Peer Review Mechanism<br>*New Partnership for Africa's Development*<br>(Indigenous) | | |
| Logical Framework<br>Analysis/Results-Based<br>Management<br>*Many bilateral development agencies as well as multilateral development banks (e.g., World Bank)*<br>(Adopted) | | |

**Table 3.1:** A more inclusive approach to capturing the developments in the field of evaluation (Carden and Alkin, 2012: 115)

Therefore, the US-centric mapping of the field of evaluation by Christie and Alkin (2013) was approached with caution in this study but, due to my focus on the English context, its conceptual tools were deemed useful in informing the conceptual framework of this research, discussed later in this chapter. Following Christie and Alkin (2013), this section now proceeds to present the foundations of evaluation theories: *social accountability*, *social inquiry* and *epistemology*, as well as the main concerns of different evaluation theorists placed on three

branches of the 'Evaluation Theory Tree' related to: *use*, *method*, and *valuing*.

### 3.2.1 Social accountability and the use of evaluation

Evaluation theorists whose work stemmed from the *foundational idea* of *social accountability* emphasised the importance of accountability in the context of the evaluation of government-funded programmes and policies. According to Christie and Alkin, the evaluator's focus on *social accountability* 'situates and legitimises evaluation as a fundamental process for generating systematic information for decision making' (Christie and Alkin, 2013: 15). In the UK context, accountability has also been cited as one of the main purposes of evaluation (gov.uk, 2020). However, accountability is typically left as an undefined, unexamined, 'underexplored concept whose meaning is evasive, whose boundaries are fuzzy, and whose internal structure is confusing' (Schedler, 1999: 13). In its simplest form, accountability refers to an 'account-giving relationship' whereby one party, an individual or institution, is under obligation to account for their actions or performance to another party (Brundrett and Rhodes, 2011). Schedler (1999) extends this basic definition and refers to two connotations of accountability: 'answerability' and 'enforcement'. According to Schedler (1999: 17):

> A is accountable to B when A is obliged to inform B about A's (past or future) actions and decisions, to justify them, and to suffer punishment in the case of eventual misconduct.

'Answerability' refers to an obligation to inform and account, in order to provide 'explanation and justification' of actions undertaken by the accountable party. Therefore, answerability involves monitoring and oversight. However, accountability may also be taken to contain elements of 'enforcement', which revolves around rewarding desirable behaviour and punishing undesirable behaviour. Because accounting actors do not just 'call into question' but also 'punish' improper behaviour (misconduct), accountable persons account for what they have done and why, and bear the consequences, including sanctions (Schedler, 1999: 17). Importantly, however, 'enforcement' is not a part of evaluation, because evaluation serves accountability only in relation to providing information for answerability, i.e., being answerable (Christie and Alkin, 2013). Accountability, therefore, involves giving account of one's actions in order to justify and explain.

The theories that emerged from the *foundational idea* of *social accountability* have been concerned with the *use* of evaluation findings to inform decision-making aimed at improving social programs and policies. The pioneering work in the *use* branch of the Evaluation Theory Tree has been conducted by Daniel Stufflebeam amongst others, who is also referred to as a

decision-oriented theorist (Christie and Alkin, 2013). Use and decision-oriented theories seek to assist key programme stakeholders who commission the evaluation. Stufflebeam's CIPP (Context, Input, Process and Product) model became one of the most well-known of these theories. The field then evolved towards utilisation-oriented theories, which aimed not only at informing decision making but also ensuring that the findings from evaluation made an impact on decision making, organisational change, and conceptual understandings of the programme (Christie and Alkin, 2013).

### 3.2.2 Systematic social inquiry and methods of evaluation

As can be seen in the Evaluation Theory Tree, some theorists grounded their work in *systematic social inquiry*, i.e., systematic study to investigate and explain human behaviours in their unique social settings, employing a variety of methods. However, the enduring question in *social inquiry* has been 'which methods are appropriate for the study of society, social groups, and social life' (Christie and Alkin, 2013: 15). Is human behaviour explainable and predictable, or is it something that can only be interpreted? While some social researchers favoured the use of statistics and experimental methods, others favoured qualitative studies of the social world. The debates regarding the most appropriate methodology to study the social world continue to this day, with impact on the practices of evaluators and the development of the field of evaluation.

Evaluation theories that stem from the *social inquiry* foundation are thus primarily focused on research *methods* enabling evaluators to obtain 'the most rigorous knowledge possible given the contextual constraints' and employ 'a methodological and justifiable set of procedures for determining accountability' in evaluation work (Christie and Alkin, 2013: 11). The fundamental works in the methods branch include publications by Donald Campbell (1957). In the UK context, the work of Ray Pawson (2013) has sought to advance the development of 'evaluation science'. Pawson problematises methods in evaluation as follows:

> Methods gain their spurs by thoughtful adaptation rather than mindless replication. Methods come and go but some are more adaptable and thus sustainable. The underlying dynamic…, it would seem, is one of the 'survival of the fruitful'. (Pawson, 2013: xii)

For Pawson and his co-author Tilley (Pawson and Tilley, 1997), these 'fruitful' methods embrace a generative view of causation, the 'challenge of complexity' and the realist assumption that, although complex, programmes, policies and social realities in general can be objectively evaluated. From this ontological foundation, Pawson and Tilley (1997) developed 'realist evaluation' that uses the CMO (Context, Mechanism, Output) model of evaluation that

recognises both the context and the complex causes and effects that contribute to behavioural change in social programmes and interventions.

The recognition of complexity in policymaking and policy implementation has also led to the development of new methodologies for policy evaluation by scholars at CECAN (Centre for the Evaluation of Complexity Across the Nexus) at the University of Surrey in the UK. These methodologies work with complexity and seek to refine methodologies for complexity-appropriate evaluation. These methodologies rely on the 'appropriateness' of the evaluation questions and methods which are co-produced with users, adaptable, iterative, and often a combination or hybrid of other methods (Befani, 2020). These methods include systems mapping, qualitative comparative analysis, theory of change mapping and broader methods guides (CECAN, 2022).

### 3.2.3 Epistemology and valuing (judging the value) the evaluand

Another *foundational idea* which has shaped the work of evaluation theorists is *epistemology*, which deals with the nature and validity of knowledge. Epistemological arguments about knowledge centre around 'the legitimacy of value claims, the nature of universal claims, and the view that truth (or fact) is what we make it to be' (Christie and Alkin, 2013: 11). Three broad epistemological stances include: post-positivism, which sees the truth as something that can be established through robust methods; constructivism, which emphasises multiple, subjectively experienced social realities, and pragmatism, which embraces both objectivity and subjectivity, depending on the context, to argue that a particular explanation of reality "may be considered 'truer' than another" (Christie and Alkin, 2013: 17-18).

The theories which grew from the foundation of *epistemology* are preoccupied with 'valuing' (value judgement) in the evaluation process. The leading work of the 'valuing' branch has been conducted by Michael Scriven (1967) and Robert Stake (1967) who argued that placing value judgement on the subject of the evaluation, *the evaluand*, is essential to the process of evaluation. The theories however split into two in the 'valuing' branch in terms of 'objectivity' and 'subjectivity' of making a judgement of the evaluand (Christie and Alkin, 2013: 12). Objectivist theorists such as Scriven, advocate that value judgment should be based on 'observable data' about the quality and effectiveness of the evaluand (Christie and Alkin, 2013: 31). They thus strive for objective judgment about the merit or worth of the evaluand. Subjectivist theorists, for instance Elliot Eisner, advocate for 'publicly observable facts' to inform the valuing process, because human elements make all value judgments subjective. In

a similar vein, Stake challenges evaluation conducted in the spirit of obtaining objective information. His position assumes and highlights the complexity inherent in evaluation. There is also a difference between theorists of the 'valuing' branch in terms of how they define a role for the evaluator. For instance, Scriven and Eisner see the evaluator as a 'valuer' who makes authoritative value judgments pertaining to 'good' and 'bad', whereas Guba and Lincoln point out there is no 'one reality' in which an object can be unequivocally judged as 'good' or 'bad'. MacDonald (1974, 1976), a UK-based evaluator, concurred with Guba and Lincoln, suggesting that evaluators need to take on the role of negotiators who acknowledge multiple views from stakeholders and seek to decrease political partiality with independent evaluators.

## 3.3 'Traditional' and 'developmental' approaches to evaluation

This section discusses another way of systematically capturing different approaches to evaluation, offered by Patton (2006) as a distinction between 'traditional' (also referred to as: classic, dominant or mainstream) and 'developmental' evaluation. Developmental evaluation, first described by Patton in 1994 and evolving ever since (Patton, 1996, 2006, 2011, 2021), emerged in the course of Patton's critique of Scriven's (1967) distinction between 'formative' (improvement-oriented) and 'summative' (judgment-oriented) evaluation. For Patton (1996), this distinction is limited, because it is not able to capture the entire array of evaluation purposes. As he emphasised, 'traditional' forms of evaluation, including 'formative' and 'summative' evaluation, miss such opportunities as: a 'conceptual use' of evaluation findings, supporting long-term improvement aimed at development, and the benefits of the process of evaluation (Patton, 1996: 131). Importantly, he argued that the developmental evaluation approach offers a range of benefits that cannot be obtained through the traditional forms of 'formative' and 'summative' evaluation. In his later works (Patton, 2006, 2011, 2012), he started to refer to 'formative' and 'summative' evaluation as 'traditional evaluation approaches' because formative and summative evaluation are very alike in their practice and purposes.

The purpose of developmental evaluation makes it distinct from traditional evaluation. Specifically, while the main 'purpose' of developmental evaluation is the 'development' of innovation and adaptation, the purposes of traditional evaluation approaches are 'improvement' (formative evaluation) and 'external accountability (summative evaluation) or getting ready for external accountability (formative evaluation)' (Patton, 1994: 318). Patton prioritised change and adaptation for development over improvement: whereas improvement

entails 'making something better' by measuring progress, a developmental evaluator provides evaluative information with 'real-time feedback' (Patton, 2006: 110) to ensure that evaluations have a lasting impact and help 'ongoing development' (Janssens and Ehren, 2016: 94). For Patton (1996: 135), traditional forms of evaluation are not sufficient to support development, long-term continuous improvement and innovation. Traditional approaches focus mainly on the expected outcomes which are a measure of improvement, but they do not take into account unexpected outcomes and innovation. In fact, unexpected outcomes or innovative approaches could be even counted as 'failure' because they do not fit within predetermined outcomes (Patton, 2006).

For Patton, traditional forms of evaluation assume that programmes are stable and fixed or 'consistent' (Patton, 2011: 23-26). They measure 'success' against clear, specific, measurable, and predetermined goals and outcomes that are assumed to be 'predictable' from the start and achieved through a 'linear cause-effect intervention' (Patton, 2006: 30). In this approach, the role of evaluators is to decide whether the programme was successful by testing the predetermined indicators of the set goals and objectives. It is important to note that formative evaluation can be 'process oriented' and not 'outcome oriented'; but for Patton (1994, 1996, 2011), many formative evaluation practices deal with predetermined 'outcomes' to make a decision about the worthiness of a 'fixed' programme in a summative manner to receive funding for a new cycle of evaluation. For Patton, traditional evaluation approaches are not interested in embracing 'complex situations' and searching for 'what is effective, for whom and under what conditions' and, as a result, may end up measuring the wrong outcomes. Measuring the wrong outcomes may, in turn, cause people to engage in wrong actions (Guijt et al., 2012: 15). Developmental evaluation is, on the other hand, driven by the idea that 'there never will be a model; the model will never be fixed' and the evaluator knows that they are evaluating 'something that keeps changing and adapting' (Guijt et al., 2012: 2). Developmental evaluation thus embraces the complexity and non-linearity of programmes (Guijt et al., 2012: 2) and values the emerging needs and context of the evaluands, reflecting those needs in the evaluation design and amending the design if necessary. Developmental evaluation thus enables evaluators and practitioners delivering programmes that are being evaluated to adapt to changes in their dynamic environments. Such changes can be about 'understanding, the characteristics of participants, technology, or the world' (Patton, 1996: 136). This means that unexpected outcomes or innovations are not counted as 'failure' but as something new that merits further consideration.

Developmental and traditional evaluation also assign different roles to evaluators. The primary role of developmental evaluators is to 'infuse team discussions with evaluative questions, thinking, and data, and to facilitate systematic data-based reflection and real-time decision-making in the developmental process' (Patton, 2012: 102). Developmental evaluators 'ask evaluative questions', 'apply evaluation logic', 'gather information and report evaluative data to support project, program, product, and/or organisational development with timely feedback' and make course corrections along the emergent path. For Patton, an evaluator is a 'facilitator and learning coach' who brings 'evaluative thinking to the table' (Guijt et al., 2012: 2; Patton, 2006: 28; Patton, 2011).

In developmental evaluation, the evaluator and the evaluand work collaboratively as partners based on a 'conversation with people around the issue of what is being developed' (Guijt et al., 2012: 18; Patton, 2006, 2011). Because developmental evaluation is relationships-driven, nobody is outside of the action, and all are a part of the evaluation. They all take part in the conceptualisation, design and testing of new approaches, as well as the 'on-going process of adaptation', 'intentional change', and 'development' (Guijt et al., 2012: 3). It is, therefore, also important to have 'developmentally oriented leaders in organisations' who are social innovators and change agents (Patton, 2006: 31).

Baseline evaluation is an important part of both traditional and developmental evaluation approaches, used to determine what information is needed and whether or not to start evaluation. However, while traditional evaluators typically assume that this early information is 'true'; developmental evaluators argue that early evaluation may not give an accurate picture of problems and may not be helpful to understand the context. A more accurate picture of the contextual needs and problems can be obtained through subsequent evaluations, once trust and relationships have been built with staff delivering a programme and the intended beneficiaries. Developmental evaluators often 'go with the flow' and observe what emerges (Guijt et al., 2012: 13), following naturalistic inquiry rather than systematic inquiry as their preferred methodological approach. The table below summarises other distinctions between traditional and developmental evaluation, some of which will be explored in the following sections.

| Evaluation Component | Traditional Evaluation | Developmental Evaluation |
|---|---|---|
| Purpose | Improvement, accountability | Development, adaptation |
| Situation | Stable, goal-oriented, predictable | Complex, dynamic, changing |
| Mindset | Effectiveness, impact, compliance | Innovation, learning |
| Target | Programme participants | Participant's environment |
| Measurement | Based on predetermined indicators | Based on emergent indicators |
| Unexpected consequences | Paid token attention | Paid serious attention |
| Evaluation Design | By evaluator | Collaborative with programme staff |
| Evaluation Methods | Based on social science criteria | Based on evaluation use criteria |
| Evaluation results | Best practices | Best principles |
| Evaluator Role | Independent from programme | Integrated with programme |
| Evaluator qualities | Strong methodological skills, credibility with external authorities and funders | Strong methodological skills, credibility with organisational and programme staff |

**Table 3.2:** A comparison of traditional and developmental evaluation approaches (adapted from Patton, 2011, cited in Fagen et al., 2011: 647)

### *3.3.1 Methods employed in traditional and developmental evaluation*

The primary methodology of traditional evaluation relies on working with predetermined outcomes and indicators and 'fixing the design upfront' with 'fixed' questions, underpinned by valuing certainty and generalised evaluation findings to disseminate (Patton, 2011: 23-26). As a consequence, traditional evaluators employ 'rigid, mechanical' designs to evaluate programmes (Guijt et al., 2012: 1). The traditional approach to accountability is to evaluate:

> whether resources are used as planned, and whether targeted outcomes are attained. This is a static and mechanical approach to accountability that assumes designers know, three or five years in advance, what important outcomes to target and how to go about achieving those desired outcomes. Departing from planned implementation is considered implementation failure. Targeting new and emergent opportunities is considered 'mission drift.' The mantra of traditional, static accountability is to plan your work, work the plan, and evaluate whether what was planned was achieved. (Patton, 2012: 110)

However, this static methodology often 'interferes' with innovation and the complexities of the context (Guijt et al., 2012: 2). Further, SMART (specific, measurable, achievable, relevant, time-bound) objectives utilised by traditional evaluators can do 'harm by limiting responsiveness and adaptability' (Patton, 2012: 104). Overall, traditional evaluation values 'standardisation of inputs', 'consistency of treatment', 'uniformity of outcomes', and 'clarity of causal linkages' (Patton, 2006: 31).

By contrast, developmental evaluation assumes a 'world of multiple causes, diversity of outcomes, inconsistency of interventions, interactive effects at every level' (Patton, 2006: 31).

Therefore, developmental evaluation is methodologically 'agnostic' and uses many methods and methodological combinations to evaluate the different parts of organisations differently, employing for instance quasi-experimental methods or interviews. It is the focus of evaluators while employing an evaluation methodology that makes it different from traditional approaches. The focus is on emergent outcomes and indicators with a 'flexible design' and 'dynamic questions' (Guijt et al., 2012: 13). This focus often requires adapting the methodology if necessary. Developmental evaluation also differs from traditional evaluation in terms of questions such as: 'what is getting developed' and 'what are the implications of the development'. It also considers 'what is not getting developed' if necessary (Guijt et al., 2012: 2). However, as Patton (2011) and Guijt et al., (2012) have noted, such methodological flexibility could provide a challenge to bureaucracies. Furthermore, few, if any, valuation designs in the public sector allow a budget for open-ended fieldwork. For financial and methodological reasons, the current social and political climate favours 'replicability, efficient goal attainment, clarity of causal specificity, and generalisability' over emergence (Guijt et al., 2012: 16).

### 3.3.2 Use in traditional and developmental evaluation

The importance attached to the use of evaluation findings emerged in parallel with a focus on *social accountability* (Christie and Alkin, 2013). Interestingly, Patton (2006) attributed accountability to the traditional evaluation approaches, specifically, accountability to external authorities and funders. This form of accountability has been used to 'control and locate blame for failures' (Patton, 2006: 30) with a 'baggage around compliance' (Guijt et al., 2012: 17). For Patton, using evaluation findings to make practitioners accountable is not the route to developing 'high-performance organisations' (Patton, 2012: 110). Evaluating organisations within a rigid, narrow view of accountability for performance data (Guijt et al., 2012: 13) means that 'new opportunities and adjustments to changing conditions will be evaluated negatively' by evaluators (Patton, 2012: 111).

In place of a 'classic accountability-oriented approach to evaluation' which embraces a 'preconceived path', Patton (2012: 111) proposed a complexity-based developmental evaluation, which embraces the 'emerging paths.' Therefore, accountability should be operationalised by documenting the 'developmental shifts.' This involves 'making the data transparent on which changes are made and documenting deviations from the original plan in terms of both implementation and the emergent outcomes' (Patton, 2012: 111). In

developmental evaluation, everyone is involved in the evaluation process, and they should be honest. Those who are being evaluated should not be afraid of documenting their failure, because they are allowed to 'justify themselves'. Patton explained in detail how complexity-based developmental evaluation reconfigures accountability as follows:

> Complexity-based developmental evaluation shifts the locus and focus of accountability. Accountability in developmental evaluation means documenting adaptations and their implications, not evaluating rigid adherence to planned implementation and preconceived outcomes. Why? Because complexity-sensitive developmental evaluation assumes that plans are fallible, based on imperfect information and assumptions that will be proven wrong, and that development occurs in dynamic contexts where even good plans will have to be adapted to changing realities. Thus, rather than becoming a barrier to adaptation, as occurs in traditional rigid accountability measures in which programmes are deemed to have failed if they depart from what was planned, developmental evaluation assumes a dynamic world with departures from initial plans. Developmental evaluation places the emphasis on understanding, supporting, and documenting adaptations and their implications. (Patton, 2012: 111)

For Patton (1996: 132), traditional forms of evaluation, including summative and formative evaluation, are only capable of generating 'instrumental' knowledge by creating a 'basis for subsequent summative decisions or program improvements.' They are however not capable of generating knowledge for a 'conceptual use' of evaluation findings. Conceptual use, contrary to instrumental use, does not focus on the immediate decision-making and taking action as a result of evaluation but instead seeks to influence the thinking of practitioners whose programmes and practices have been evaluated. Patton (1996: 134) argued for an evaluation approach which focuses on knowledge generation to 'learn' from the programme's effectiveness and to understand 'what works', rather than conducting evaluation to measure whether the programme has worked and to 'judge the merit or worth of individual programs.' Developmental evaluation supports 'hunger for learning' (Patton, 2006: 30).

Patton (1996, 2011) was also critical of traditional evaluation approaches as they did not concentrate on the uses of the evaluation processes itself, because 'use' was framed in these approaches by questions of 'what happens to findings and recommendations' (Guijt et al., 2012: 3). Patton's (1978, 1986) earlier work, titled 'utilisation focused evaluation', identified the primary users and clarified the intended uses of evaluation findings. Over time, Patton (1997, 2006, 2011) also noticed the importance of 'process use' rather than solely focusing on using evaluation findings. Process use refers to changes in behaviours, actions, and thinking of individuals which make a difference in the evaluand's practice through the 'learning that occurs during the evaluation process'. Such changes consist of both individual and organisational

changes (Guijt et al., 2012: 3). For Patton, process use is a missed opportunity in traditional forms of evaluation because 'much of the impact of evaluation comes from the process of engaging in the evaluation, not from the findings' (Patton, 1996: 137). Process use helps in two ways: by supporting interventions and by empowering participants (Patton, 1996).

### 3.3.3 Valuing the evaluand in traditional and developmental evaluation

As explained in Section 3.2 above, 'valuing' (judging) is an essential element of evaluation. For instance, for Scriven (1965), there is no evaluation until judgment has been passed, and the evaluator is the best-qualified person to judge. Some theorists, for instance, Suchman (1967) argue that valuing is what makes evaluation different from research. However, for Patton, looking for 'definitive judgments of success or failure is very much the work of traditional evaluation', with its overarching goals of 'high fidelity' in valuing and 'generalisable findings across time and space' (Patton, 2006: 30). The valuing occurs in comparing the outcomes with predetermined indicators and goals. The negative consequences of external evaluation, or 'intrusive external judgment', can be 'painful' in many contexts (Guijt et al., 2012).

By contrast, developmental evaluators see valuing differently. Rather than drawing conclusions on the merits and worth of programme to provide a definitive judgement, they seek to provide feedback that is nuanced and disaggregated in order to 'generate learning', 'support direction' or 'affirm changes in direction' (Patton, 2006: 30). Developmental evaluation does not ignore the contextual information and aims to produce 'context-specific understandings' that inform ongoing innovation and adaptation (Patton, 2006: 30; Patton, 2011). Therefore, developmental evaluators use the findings of evaluation to identify 'what works for whom in what ways under what conditions' (Guijt et al., 2012: 13).

### 3.3.4 Values in traditional and developmental evaluation

'Evaluation' is about values and valuing, but they are not the same. Whilst valuing might refer to rationally determined conclusions (Christie and Alkin, 2008), values refer to 'beliefs, preferences, interests, wants, needs, or desires' (House and Howe, 1999: 6). The process of valuing is critical and often constitutive of a hierarchy of values of the evaluator. Values can be both individual and social. Collective values in a given society or a public education system may inform some evaluation approaches such as 'democratic evaluation', an approach that is designed and carried out in the interest of the public (MacDonald, cited in Norris, 2015). However, values have been subsumed in Christie and Alkin's (2013) Evaluation Theory Tree under the branch of valuing. It could be argued that this is a reductive perspective, and values

would merit a separate branch, with a primary focus on questions related to values, such as: 'which values' and 'whose values'. The work of UK-based theorists (MacDonald, 1974, 1976; Norris, 2015; Picciotto, 2015; Simons, 1987) can be seen as based on a *foundational idea* (Christie and Alkin, 2013) of democratic values (Section 3.4).

In traditional evaluation, it is the values of external evaluators (or those who commission evaluation) that determine the outcomes and indicators; as suggested by Patton (2006, 2011). The evaluator controls the evaluation and the evaluator's perspective, and values are imposed on the evaluands. This is related to the 'control' function of accountability, and it is the evaluators (or their employer) who control evaluation (Guijt et al., 2012: 13) to assure that the values of the 'top' reflect the evaluation practices. This approach positions the evaluator 'outside' to assure the independence and objectivity of evaluation (Patton, 2006: 30).

In developmental evaluation, on the other hand, the values of the evaluand are as important as those of the evaluator and the aim is to arrive at some shared values. To achieve this, the evaluator collaborates to 'design a process that matches philosophically and organisationally' (Patton, 2006: 30). Accountability in developmental evaluation is thus centred on the evaluand's deep sense of 'fundamental values and commitments' (Patton, 2006: 30). Due to collaboration and the sharing of values, in developmental evaluation, the external evaluator cannot play the role of an 'independent external evaluator' (Guijt et al., 2012: 18). Developmental evaluators are integrated into ongoing interpretative processes (Patton, 2006: 30) and consequently, should have a flexible outlook:

> Not every personality type resonates with the role of developmental evaluation evaluator. Obsessive types with rigid personalities and a high need for control, do not make good developmental evaluation evaluators. If you aspire to be a developmental evaluator, you need to examine yourself: it is fit for you, can you thrive in that climate, and can you play a supportive role for people? (Guijt et al., 2012: 17)

Importantly, developmental evaluation sees the external evaluators as vital for ensuring rigour, because 'somebody internally involved in developmental evaluation is not in the position of assessing the merits and worth of a programme' (Guijt et al., 2012: 17). The position of external evaluators has been further elaborated by MacDonald (1974, 1976), who highlighted the importance of democratic evaluation.

## 3.4 Democratic evaluation

Barry MacDonald's (1974, 1976) work on democratic evaluation could be read as stemming from an assumption that evaluators who practise within the public domain cannot escape the

world of politics. In the words of Helen Simons, who also sought to advance this approach in the UK context, MacDonald's work was a response to a growing awareness of the need to shift power in research relationships 'away from regarding the evaluator as the sole judge of what was worthwhile' (Simons, 2009: 17). The advocates of democratic evaluation of school curriculum reforms in the UK were also concerned with 'finding ways of getting to know schools that will enable them to become more democratic institutions offering a more educational service' (Simons, 1987: 1).

In order to systematically capture the importance of democratic evaluation, MacDonald (1974, 1976) distinguished between three main orientations prevalent in the evaluation of curriculum reforms and social policies at the time: 'democratic', 'autocratic', and 'bureaucratic' evaluation. He argued that evaluation is a political activity, and no evaluation is value-free. Any evaluation design, with its selection of approach, roles, goals, audiences, issues, and techniques, is affected by diverse values and interests. Therefore, the information generated by evaluation is a 'resource for the promotion of particular interests and values' (MacDonald, 1974: 132). For instance, bureaucratic evaluation is driven by the values of the contractual obligations (i.e., of government), and the evaluator's room for manoeuvre is restricted by the contractual obligations. Autocratic evaluation, such as evidence-based evaluation, which thrived since the 1980s within the premises of NPM, takes a value-free stance by 'clothing it in technocratic apparel' and giving 'pride of place to the achievement of pre-determined goals' and 'thriving on tracking progress through indicators' (Picciotto, 2015: 156). However, for MacDonald (1974, 1976) and other democratically oriented evaluation scholars, evaluation should be driven by democratic values, i.e., by 'acknowledging the perspectives of participants in evaluation process judgements of those who have decisions to make in policy or practice context' (Simons, 2009: 17). For instance, Howe (2003) referred to the central place of values in educational research as follows:

> Educational research [and evaluation] can never be value-free. To the extent it approaches value-freedom in its self-perception, it is to that extent dangerous… [and] in fact… useless… I take it as a given that democratic values are prominent among those that educational research [and evaluation] ought to incorporate, a premise not likely to be challenged in the abstract. (Howe, 2003: 133-134)

MacDonald (1974, 1976) concentrated on the relationships between evaluators and programme participants (including other stakeholders) and argued that democratic evaluators should be independent to serve the 'whole community' by taking on a 'brokering' role. They should be 'explicitly recognising and valuing pluralism' by seeking to 'represent a range of interests and

values in their work' (Norris, 2015: 136). Within that, evaluators should engage in 'periodic negotiation' of their relationships with sponsors and programme participants (MacDonald, 1974: 134) pertaining to the evaluation focus, content, and reporting. Democratic evaluation should foster participation and collaboration by including stakeholder's voices, to serve the whole community as an information service. Therefore, the success of evaluation depends on the 'range of audiences served' (MacDonald, 1974: 134). In democratic evaluation, the relationship between the evaluator and informant is built on 'confidentiality' as an ethical responsibility. For Norris (2015: 136), these important principles minimise the 'threats and risks those individuals might feel because of external evaluation' and maximise participation (Norris, 2015: 136). The evaluation report should be non-recommendatory because the evaluator's role is first of all to help 'all our peoples to choose between alternative societies' (MacDonald, 1979: 12). They should also be understandable and accessible not only for those with technical background but also audiences who lack technical expertise, based on the public's right to know (Norris, 2015; Picciotto, 2015). In this regard, evaluation should be an 'impartial information service to all stakeholders, participants and the public on the value of the policy or programme to enable them to contribute to informed policy-making and debate' (Simons, 2006: 245-246). For Simons, evaluators' freedom from particular interests is essential to 'be fair and offer a sound basis for informed action' (2006: 245-246).

However, for other evaluation scholars, MacDonald's democratic evaluation was 'too idealistic' (Picciotto, 2015: 159). For instance, House and Howe (1999) modified MacDonald's notion of democratic evaluation by calling it 'deliberative democratic evaluation' and noted that a democratic evaluator is 'not a passive bystander, an innocent facilitator, or a philosopher king who makes decisions for others, but rather a conscientious professional who adheres to carefully considered principles' (House and Howe, 1999: 111). More recently, Picciotto (2015) problematised both the independence of evaluators due to fee dependency and 'the neutral brokerage role' associated with democratic evaluation. Given its neutral procedural stance, democratic evaluation lacks 'the leverage needed to induce responsibility of authority in environments where power asymmetries hinder evaluation independence' (Picciotto, 2015: 161). He thus proposed a 'progressive (contemporary democratic) evaluation approach' as an update of MacDonald's democratic evaluation and House and Howe's deliberative evaluation models. He proposed a 'more activist and independent evaluation model grounded in professional autonomy' (Picciotto, 2015: 164) by evaluators who are funded by novel funding and sponsorship arrangements in order to break the traditional chains of fee dependence (e.g.,

independent think tanks, private donations, crowdfunding). For Picciotto, fee independence is essential to protect the independence of evaluators and the integrity of evaluation processes. He supported his argument with an example of 'undemocratic and illiberal regimes whether market oriented or state driven or both' in the 21st century, which prevent evaluators from engaging in the 'logic of dialogue, inclusion and deliberation' (Picciotto, 2015: 161). As he put it:

> It is not enough for democratic evaluators to provide neutral information services, broker debate and facilitate deliberative decision-making processes. The time has come to experiment with a more activist and independent evaluation model grounded in professional autonomy, reliant on independent funding sources and tailor made to diverse governance environments. Such a model would be designed to favour social equity, level the playing field of decision making and break the chains of fee dependence. It would be grounded in moral ethical precepts, own evaluation products and engage in the policy-making process through alliances with progressive forces. It might even pave the way for a democratic wave of evaluation diffusion in the 21st century.
>
> Thus, progressive democratic evaluation would assert its independence in shaping evaluation agendas and selecting evaluation methods. It would emphasise compliance with ethical standards, professional autonomy, analytical rigor and engagement with citizens and the civil society. It would reject the single technocratic and minimalist narrative about the role of evaluation in society currently associated with value free, narrowly conceived conceptions of experimentalism. It would embrace mixed methods, pluralistic modes of inquiry and inclusive evaluative processes. (Picciotto, 2015: 161-164)

Overall, despite some differences, democratic evaluation theorists value both the independence of evaluators and democratic values in evaluation process. In this regard, Norris (2015: 138) highlighted the benefits of democratic evaluation as follows:

> it is especially suited to programmes which are strikingly new or innovative and likely to generate strong and contradictory reactions, to programmes that are surrounded by conflicts of value, interest and ideas, to programmes where hierarchy and territory in the organisation make it difficult to understand, plan and to implement change and programmes that are highly politicised and where there is little or no agreement on the facts of the matter or the kind of information that is relevant for decision making.

Importantly, democratically oriented evaluation approaches do not emphasise 'prescriptions about the technical aspects of evaluation practice' but focus on the 'relationships evaluators establish with others in a given context and the processes and interactions that enact these relationships' (Greene, 2006: 119). In this regard, Norris (2015: 138) argues that a democratic evaluator resists methodological dogmas, i.e., a conviction that particular research designs are 'intrinsically good', irrespective of the complexities of the context and plurality of values.

## 3.5 A conceptual framework for rethinking school evaluation in England

This section draws on the discussions above to develop a conceptual framework for rethinking school evaluation in England. As is apparent from discussions in Chapter 2 and Section 3.3 of the present chapter, the official approach to school evaluation approach in England since 1988 has been largely informed by what Patton (2006, 2011, 2012) referred to as the 'traditional' approach to evaluation, which constructs the purpose, method, use and valuing in particular ways. This official approach appears to be both limited, when considered in the light of the work of evaluation theorists such as Patton (2006, 2011, 2012) and limiting, when its negative consequences reported in Chapter 2 are considered.

Section 3.5.1 focuses on rethinking the purpose of school evaluation, informed by concepts from developmental evaluation (Patton, 2006, 2011, 2012). The key point here is that schools should not be left alone and can develop by learning from other schools, with collaborative learning as the main purpose of school evaluation. Although the benefits of developmental evaluation over traditional evaluation approaches have been noted time and again by Patton in his various publications (Patton, 1994, 1996, 2006, 2011), developmental evaluation is not a 'default option' to employ, and there are situations in which it may not be suitable or even risky to use (Guijt et al., 2012: 11; Patton, 2011). Therefore, elements from traditional evaluation approaches continue to be of benefit because the needs of a government with a social purpose include public accountability, in the sense of 'answerability' (Schedler, 1999) discussed above. I then move on to discuss the implications of traditional and developmental evaluation approaches within school evaluation settings and draw on Christie and Alkin (2013) to rethink methods, use, and valuing in school evaluation. Section 3.6 proposes a new set of *pillars* for school evaluation in England that draw both on the discussions presented in this chapter and the critiques of the official approach presented in Chapter 2. These *pillars* include collaboration by trusted professionals, whose voices matter and whose voices are heard, and capacity building, resting on the *foundational idea* of democracy and recognising schools as complex organisations which operate within specific contexts.

### 3.5.1 Rethinking the purpose of school evaluation

As discussed in Chapter 2, the main purposes of school evaluation in England since 1988 have revolved around school 'improvement' and 'accountability', and those purposes sit within the traditional approaches to evaluation (Patton, 2006, 2011). However, there is tension between

accountability and improvement (Ehren, 2016). For McNamara and O'Hara, 'in an ideal world' school evaluation might be able to deliver on all policymakers' goals simultaneously but, in practice different goals are likely to be 'competing' with each other rather than being 'complementary' (McNamara and O'Hara, 2009: 275). When schools are accountable for centrally defined improvement, it is hard for them to be open and honest about their weaknesses. School evaluation may become a tool for government to control schools and school improvement may be limited to externally set criteria that fail to account for political, social, economic, and contextual factors of individual schools. For instance, since its inception in 1992, Ofsted's frameworks explicitly emphasised school improvement as the overarching purpose of school inspections. The role of inspectors was referred to as 'diagnosis' in the 2015 Ofsted Framework, and in 2019, this was changed to 'identifying' what schools should improve, within standardised frameworks. This improvement-oriented purpose of Ofsted's evaluation of schools, is aligned to traditional evaluation. However, aiming to improve through evaluation assumes that 'what was done before' by schools 'was inadequate or less effective' (Patton, 1996: 135) and, consequently evaluation findings can be only used 'instrumentally' to comply with inspectors' recommendations. This make take the ownership of improvement away from schools and their local context.

By contrast, employing developmental evaluation is suitable for 'programmes undergoing continual development and change, where goals to standardise a program model and test its effectiveness may be completely absent' (Dupuis et al., 2013: 258). Schools are complex organisations, and a school is never 'stable' and never only delivering 'consistent' or 'fixed' programmes of study (Patton, 2011: 23-26). Therefore, schools should not be evaluated as if they simply aimed at preconceived outcomes, but rather as systems that are alive, evolving, faced with different needs and constantly adapting to their changing environment. Consequently, schools would benefit the most if evaluated in alignment with developmental evaluation. The purpose of school evaluation should not be to measure their improvement according to rigid, standardised criteria, but rather according to how they develop the different aspects of their provision as they unfold in non-linear ways. Enacting traditional evaluation has often been done in the name of giving students equal opportunities. However, the linear approaches of traditional evaluation that rely on measuring inputs and outcomes, through pupil performance data and Ofsted grades may have detrimental impact on equal opportunities (Patton, 2012).

Developmental evaluation is based on the idea that 'there never will be a model; the model will never be fixed; it is constantly adapting' (Guijt et al., 2012: 2). A starting point in developmental evaluation would be acknowledging that schools are learning organisations and they can best serve their students if they focus on collaborative development. For a collaborative development of schools, evaluators should act in a partnership relation and evaluation should inform and support innovation and adaptation (Patton, 2011). Evaluating in this way would support schools in learning to think evaluatively and adapt their existing practices and internal policies to their complex and rapidly changing environments. Because the values of schools can be reflected in the evaluation design, this would encourage schools to describe what 'development' and 'improvement' means in their own individual context. They could be more 'creative' and 'innovative' without fear of fitting into the external requirements. Then schools could focus on their current and emergent needs. For instance, in disadvantaged schools whose students face domestic violence, teachers would prioritise children's rights and emotional support by collaborating with social services and appropriate specialist support. There is not much literature about the implications of developmental evaluation in a school context but Janssens and Ehren (2016: 94) argued that developmental evaluation may be a particularly good fit for school networks that have some element of innovation in their vision, because it informs the future directions for a network.

Whereas developmental evaluation would give schools more freedom to shape their own improvement, according to the specific needs of their students and the local context, one element of traditional evaluation is important: as explained in the 2015 and 2019 Ofsted Frameworks, inspectors' evaluation of schools serves the purpose of providing 'assurance to the public and government that minimum standards of education, skills, and childcare are being met; that public money is being spent well; and that safeguarding arrangements are effective' (Ofsted, 2015: 9; Ofsted, 2019a). This purpose of evaluation is linked to 'external accountability' and schools in England are also accountable to parents. Accountability measures that are based on performance data narrow the benefits that schools could gain from the evaluation process and seem to be serving the needs of the Government, who need to check how public money is utilised for school improvement. The question, therefore, is not so much about what should be measured in the evaluation of schools or how to measure school improvement. Rather it is about how to deal with the tension between collaborative developmental needs of schools and the risk management needs of the government. If we consider development (through collaborative learning) as a core purpose of school evaluation,

risk management could be considered as another core purpose of evaluation. These two purposes could be seen as 'twin' purposes of school evaluation. Whilst developmental evaluation serves the needs of the school, it cannot meet the needs of governments as it is unable to directly demonstrate that public money is well used. Nor can it offer an overall picture of the country's education system because it does not rely on national standards. This is where the idea of accountability, in the sense of 'answerability', i.e., providing 'explanation and justification' of the school's actions (Schedler, 1999) could the utilised, together with risk assessment by external evaluators such as Ofsted, characteristic of traditional evaluation.

### 3.5.2 Rethinking 'method' in school evaluation

As discussed in Chapter 2, since 1988, school evaluation in England has relied on a 'one-size fits all' methodology by valuing predetermined outcomes and expectations. This might be because test-based methodology for ranking schools is relatively easy and cheap to administer, as opposed to school self-evaluation. School league tables do not take into account 'political tides' (Shadish and Luellen, 2005: 186) which, under more bureaucratic or autocratic administrations focus on narrow performance outcomes, leading to schools losing sight of their democratic mission. Similarly, when parents are constructed as 'customers' in a marketised education system, collecting their views as part of an Ofsted inspection may be tokenistic. Whilst national standards are the same for every school, schools are complex organisations, and it is the dialogue between evaluators and schools that should be able to decide which methods to follow in order to understand schools and to help their development.

Employing developmental evaluation can 'help those involved in or leading innovative efforts incorporate rigorous evaluation into their dialogic and decision-making process as a way of being mindful about and monitoring what is emerging' (Patton, 2006: 30). Evaluators should understand the 'social dynamics' of schools because the 'interaction of small effects' at schools can lead to patterns that are not predictable or controllable beforehand (Guijt et al., 2012: 6). Therefore, evaluation methods, procedures, and techniques should take into account the unique position of the school, at a particular time in its development. What is appropriate for one school may be less appropriate for another. Even what is appropriate for a school at one time, may not be appropriate at other times.

### 3.5.3 Rethinking 'use' in school evaluation

Within the official approach, Ofsted reports are purported to enable parental choice, with parents constructed as 'customers.' For example, the Ofsted 2015 and 2019 Inspection

Frameworks refer to the users of evaluation findings as parents, carers, learners, and employers. Inspection reports provide 'informed choices' for parents:

> The inspection provides important information to parents, carers, learners, and employers about the quality of education, training, and care being provided. These groups should be able to make informed choices based on the information published in inspection reports. Under the common inspection framework, readers will be able to compare different inspection reports quickly and easily, particularly where the reports are about provisions for children or learners of similar ages. (Ofsted, 2015: 7)

Evaluation is costly, both in terms of finances and time. It is, therefore, important to ensure that schools utilise evaluation. To best utilise evaluation, three points should be noted. Firstly, the use of evaluation should not be limited to the use of evaluation findings and should be extended to the process use (Patton, 1996). To focus on the process use, the relation between evaluators and schools should be reconceptualised and they should work as partners. The process of evaluation enables learning from the evaluation process when the interactions between evaluators and schools are based on partnership. Schools can get more benefit from developmental evaluation compared to traditional evaluation because it frames the role of evaluators in terms of reflecting on and questioning an evolving process.

The main barrier to the process use of evaluation is 'rigid accountability', which amounts to a checking exercise based on 'what is intended to do in the beginning' (Guijt et al., 2012: 10). This builds barriers between school evaluators and schools. Schools may not be very open, and school evaluators may not be interested in the contextual features of schools because they need to check whether schools achieve the expected indicators, which were identified in the beginning, from 'above'. Therefore, process use is only possible if evaluation is conducted to support the development of schools. This also requires establishing trust-based relations between evaluators and schools, as well as abandoning a hierarchical relation between the evaluator and evaluand. Under these conditions, schools can be honest rather than fabricating documents to demonstrate how good they are. Being honest about the problems they are experiencing would enable schools to find solutions, together with evaluators. This is why, accountability should also be reconceptualised. Accountability, as suggested by Patton (2011, 2012), can focus on realistically 'documenting' the development and struggles of schools. The enforcement element of accountability to 'punish' school should also be removed. Paradoxically, it is 'failure' rather than 'success' that is the seed of development in developmental evaluation (Guijt et al., 2012: vi). Public accountability, within developmental evaluation, could be based on the performance data of local authorities rather than individual

schools competing against one another. Rather than blaming individual schools or headteachers, public accountability apportions responsibility to all stakeholders, including policymakers.

Secondly, within the current official approach, schools are not considered to be 'users' of school evaluation in England. As noted in Chapter 2, even school self-evaluation has been carried out by schools based on Ofsted standards. This means that both school evaluation by Ofsted and self-evaluation serves the needs of 'national' accountability rather than schools' needs to learn and develop. In a scenario of employing developmental evaluation in English schools, the developmental evaluator would 'offer a view' that could 'strengthen the internal commitment' of schools to the evaluation findings by 'finding out what is working or not' (Guijt et al., 2012: 1). Therefore, schools could take ownership of evaluation findings; evaluation would be done *with* them rather than *to* them. As Patton explains, 'helping people learn to think evaluatively can make a more enduring impact from an evaluation than the use of specific findings generated in that same evaluation' (2006: 28). Developing a school's evaluation capacity can have an ongoing and lasting impact on 'how they think, on their openness to reality-testing, on how they view the things they do, and on their capacity to engage in innovative processes' (Patton, 2006: 28). Whilst accurate and detailed evaluation data are important for decision-making in schools, of more importance is whether schools can 'interpret' such data, make the data meaningful and use it to 'set direction and motivation in the face of ambiguities and conflicting demands' (Patton, 2006: 30).

Thirdly, a distinction should be made between 'conceptual' versus 'instrumental' use of evaluation findings (Patton, 1996: 132). Evaluation findings should be considered for their 'conceptual' rather than 'instrumental' use, in other words the findings should be recommendary. For the English system, this would mean abandoning the current 'instrumental' use of inspection findings as legally binding for schools, as clarified in the Ofsted (2015) as follows:

> Inspectors will assess the extent to which the school or provider complies with relevant legal duties (2015: 3) and will evaluate in line with frameworks, national standards, or regulatory requirements (2015: 6,17) by upholding and demonstrating Ofsted values all the time (2015: 17).

### 3.5.4 Rethinking 'valuing' in school evaluation

In the school context, the valuing (judging) branch of evaluation found reflection in Ofsted inspectors' judgment of schools expressed in terms of school grades. For Patton (2006) and

Guijt et al. (2012: 1), Ofsted grades would amount to 'intrusive external judgment'. It seems that Ofsted are aware of the negative consequences of 'judgement', as implicit in the recommendations to communicate judgement 'sensitively' and 'clearly' (Ofsted 2015, 2019a). Removing blame and judgment from the process of evaluating schools and replacing these with a system that is feedback- and learning-oriented would offer a more effective approach (Patton, 2006: 30): judgment free, blame-free evaluation can 'free' people from hiding the 'cold reality'. Judgment-free evaluation would, therefore, seem to be the most effective way of gathering accurate information about schools which would also enable evaluators to work *with* schools (Guijt et al., 2012: 2).

Also, assigning league table places to schools in the order of their achievement in high stakes tests is another way of valuing. Knowing a school's ranking within its area encourages competition rather than collaborative development. In developmental evaluation, school league tables would be considered a poor substitute for more complex, contextualised evaluation findings disseminated to the wider audience.

## 3.6 Rethinking the *pillars* of school evaluation in England
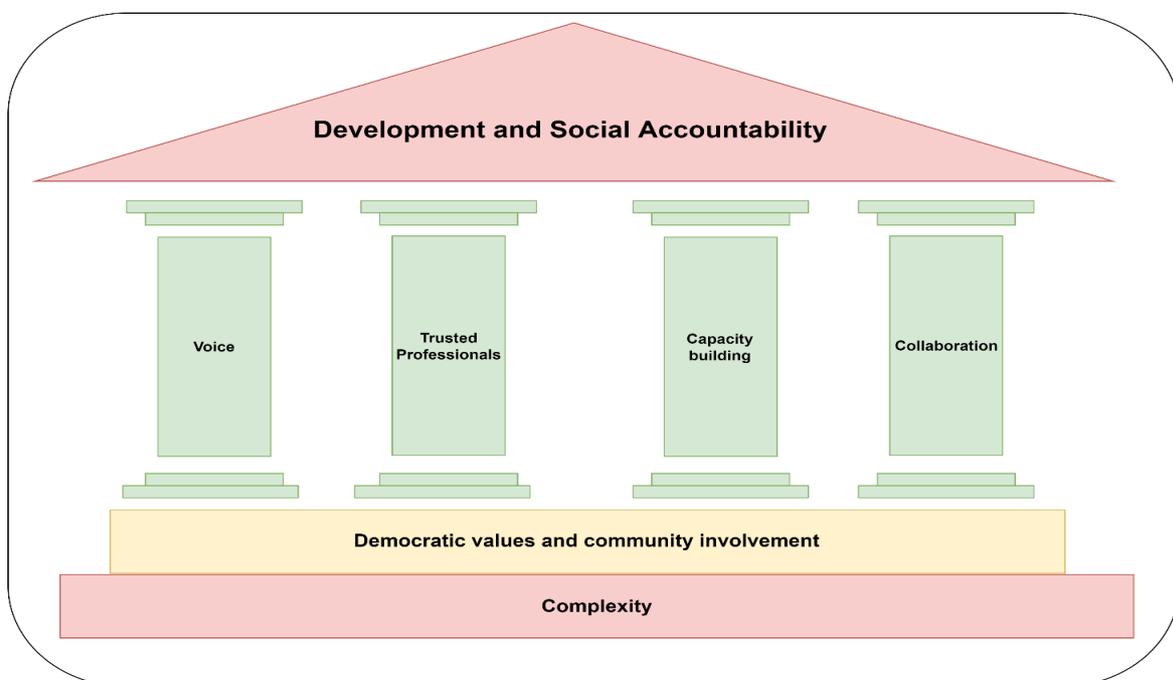


**Figure 3.2:** Alternative *pillars* of school evaluation

59

Figure 3.2 above reflects what I propose as alternative *pillars* of school evaluation. As discussed in Chapter 2, the four *pillars* of the official approach to school evaluation in England include: accountability, Ofsted inspection, school league tables, and parental choice, which are broadly aligned to what Patton (1996, 2006, 2011, 2012) has identified as a traditional approach to evaluation. The alternative *pillars*, by contrast, draw mainly on the concepts from developmental and democratic evaluation. This section will briefly discuss each of these in turn.

### 3.6.1 Voice

As explained in Chapter 2, one of the four *pillars* of the official approach to school evaluation in England is parental choice. In its place, in alignment with developmental evaluation, I would propose instead to see school evaluation as foregrounding voice, or even voices (plural), of teachers, parents, students and the local community. Although choice was equated with democracy by some writers (see Box et al., 2001), voice is more relevant to democratic relations. In its essence, parental choice is the central element in a marketised education system (Cousins and Ryan, 2009; Gewirtz et al., 1994; Ryan and Feller, 2009), but this is choice as a 'consumer' or 'user' of education, rather than a participant whose voice matters in decision-making about schools and children's education. I would argue that the idea of choice diminishes parents' role because it constructs them as 'outsiders' - the people schools should give account to. In developmental evaluation, the voices of the evaluators and evaluands need to be articulated and heard and these voices are assumed to be used with honesty. Voices of the wider community should also be included in the evaluation of schools because schools do not exist in a vacuum and benefit the local community and, in return, need local community support.

### 3.6.2 Trusted Professionals

Trust is missing both as a *pillar* in the official approach to school evaluation in England and in the political 'climate' on NPM that provided *foundational ideas* (Christie and Alkin, 2013) to evaluating schools post 1988. As Power (1997, cited in Norris and Kushner, 2007: 2) explained, NPM undermined the professions previously considered to be 'trusted':

> much of the success of the NPM has been the consensus to break the hegemony of the professions, to discipline them with market competition, and to use that competition as a surrogate for public accountability.

Just as the professionalism of teachers was diminished, the power of the central government increased, with the de-professionalised teachers 'audited' and held to account by external

evaluators - Ofsted inspectors (Ozga, 1995: 30). External evaluation based on rigid criteria forces schools to comply with what is expected and 'fabricate' evidence of compliance if necessary (Perryman, 2009: 628). Where trust is lost, teachers may 'game' the system and withhold honest responses to the evaluators, as they do not trust them. Another serious consequence of the loss of trust is that professionals may not even trust 'in themselves'. For instance, Richardson (2022) concludes that schoolteachers are not trusting in their own professional judgement about their assessment of students.

And yet, as argued by Norris (2007: 139), 'the credibility and utility of evaluation rest on trust'. In other words, collecting valid evidence from professionals requires trust and the lack of trust is a major barrier to evaluation. This is why, for more credible and useful evaluation findings, professionals should be trusted as the best people to decide what needs to improve and what they need to develop based on their own values. Also, generating accurate information without trust is impossible or very difficult. Trust is vital in every stage of evaluation:

> Evaluation stands in an interesting relationship to trust. The credibility and utility of evaluation rests on trust. Loss or lack of trust is a major impetus to evaluation, and evaluation sometimes takes the place of trust. The process of evaluation requires trust, and evaluation is used to underpin or provide a warrant for trust. (Norris 2007: 139)

Therefore, as a *pillar* of school evaluation, trust in teachers as 'social innovators' who what to do good would improve the evaluation of schools. Generating high-quality evaluation is extremely difficult in low-trust societies, the problem here is with telling the truth, especially where there are high stakes attached to evaluation. As Patton (2006) pointed out, traditional evaluation approaches are the enemy of high-quality evaluation: when schools are punished for the results which are not always in their control, but rather a result of a complex mix of their economic, financial, and political context, they may 'fabricate' information.

By contrast, developmental evaluation nurtures trust by seeing shared values between the evaluators and evaluands. Instead of trying to impose their values, evaluators work with schools to understand their values and what development and improvement may look like for them. But evaluators can still ask 'tough' questions to 'test' their beliefs (Guijt et al., 2012: 18). For this kind of relation, trust plays an essential role. Professionals need to be trusted that they want to do good.

### 3.6.3 Capacity building

Although a vast majority of literature associates capacity building with internal school evaluation, framed as professional development which aims at sustainable school improvement (see Blok et al., 2008; Nelson et al., 2015; Nevo, 2009), capacity building is also an essential element of developmental evaluation, regardless of whether it is internally or externally conducted (Patton, 2015: 14). Capacity building within the context of developmental evaluation refers to increasing the school's capabilities for evaluation, adaptation, and reflexivity (Patton, 2015). Capacity building has a dual function. The first is that evaluative, reflexive, and adaptive capabilities of schools can be an output of developmental evaluation. The second is that capacity in relation to evaluation, reflexivity, and adaptation is the condition for effective developmental evaluation. These two functions are related to the active participation of schools in the evaluation process and close working relationships with the evaluators. Within this, a developmental evaluator brings 'an evaluation perspective and evaluative thinking to the team' (Patton, 2015: 12) and schools can develop the 'capacity to do evaluation', as well as the 'capacity to use it' (Levin-Rozalis et al., 2009: 192). When school leaders and teachers have evaluation capacity, they can evaluate and adapt their own practice, reflect on their success as well as failure and use this to inform the next steps (Patton, 2015). Therefore, capacity building is essential in school evaluation for development, innovation, and change planned by schools themselves (Guijt et al., 2012: 3). Some empirical studies evidenced how partnership-based school evaluation practices (i.e., school to school peer review) enhanced schools' evaluation capacity (Matthews and Headon, 2015; Ozsezer, 2016). On the other hand, Glasswell and Ryan (2017) found that there is tension between mandatory standards documents (i.e., Ofsted school inspection framework) and schools' capacity building to evaluate themselves. This is partly because of the restrictive nature of mandatory standards:

> while the standards promote critical reflection, they rarely suggest that teachers be reflective about the larger sociopolitical aspects of schooling and education systems. (Glasswell and Ryan, 2017: 3)

### 3.6.4 Collaboration

Developmental evaluation foregrounds the collaboration between the evaluators and schools. This could help schools to develop evaluation skills, including observation, evidence gathering, and evaluation design, which would in turn strengthen teachers' confidence to contribute to the evaluation and increase the quality of evaluation. The *pillar* of collaboration is proposed here as an alternative to the idea of competition as a driver of school improvement, central in the official approach to school evaluation in England (see Chapter 2, Figure 2.1). A key problem

with competition as a driver of improvement, e.g., schools competing against one another in school league tables, is that it frames school improvement in terms of improving a single school. The premise is that competition improves, however, competition also creates winners and losers. Thus, some schools win at the expense of other schools. Therefore, competition is not a good formula for improving the education system, at the national level. For this, schools need to collaborate with one another, guided by the common purpose of improving all schools; visiting other schools and peer learning would help them to develop new strategies, strengthening their professionalism (Matthews and Headon, 2015; Ozsezer, 2016). Schools in an area (i.e., local authority) should work as partners. The wider community also should be encouraged to involve in such evaluation collaboration.

## 3.7 Conclusion

Based on the developments in the field of evaluation discussed above, this chapter drew on concepts from developmental evaluation (Patton, 1996, 2005, 2011, 2012) and democratic evaluation (MacDonald, 1974, 1976) to propose an alternative set of *pillars* for evaluating schools in England. Developmental evaluation accepts the challenges of complexity and embraces 'failure' recasting it as the 'seeds for successful innovation' (Guijt et al., 2012: vi). The developmental evaluation approach also offers concepts and tools that are absent from traditional approaches, such as: 'conceptual use' of evaluation findings; supporting 'long term ongoing improvement' and benefiting from the process use of evaluation (Patton, 1996: 132-135). Developmental evaluation can 'nurture developmental, emergent, innovative, and transformative processes' (Patton, 2006: 28). It points to an alternative purpose for school evaluation: collaborative learning.

However, this does not mean that developmental evaluation should be rigidly followed. Its main limitation in the context of evaluating schools that form a national system of education is that there is a need for national standards and risk management on the part of the Government and its agencies. As Patton (2011) points out, in risk situations, there is a need for independent evaluators. Rather than closely scrutinising professionals' work, independent evaluators would need to ensure students' safety, safeguarding and access to education. Therefore, some concepts and tools from traditional evaluation are also important in order to minimise and manage risk.

# CHAPTER 4: RESEARCH DESIGN AND METHODOLOGY

## 4.1 Introduction

This study employed expert interviews to collect data and address research questions 1-3. As explained in Chapter 1, the research questions focused on educational experts' views on the official approach to school evaluation in England (research question 1) and their suggested alternatives (research question 2). The third research question has sought to ascertain how the suggested alternatives could be assessed in light of current knowledge in the field of evaluation.

This chapter discusses the methodology developed for this research. I begin by explaining my epistemological and ontological stance and my understanding of a given approach to school evaluation as a social construct (Section 4.2). I then discuss the expert interview and how this method of data collection was deemed appropriate for collecting data that would answer my research questions (Section 4.3). Next, I describe the process of planning and conducting expert interviews (Section 4.4) with the strategies I developed to overcome challenges that stemmed from ethical issues, gaining access to the experts and interviewing with them. I then explain how I organised and analysed my data and presented my findings (Section 4.5). I then discuss ethical considerations during and after data collection (Section 4.6) and finally the validity, reliability, and generalisability of my research (Section 4.7).

## 4.2 An 'approach to school evaluation' as a social construct

The various approaches to school evaluation in England explored in Chapter 2 do not appear to have been derived from some essential, universal, and inevitable nature of education. The question could therefore be asked: 'where do these approaches come from?' They could be seen as developed (or 'constructed') by government officials and experts, civil servants working in the Department for Education, Ofsted officials and other powerful 'policy actors' (Ball et al., 2012: 49). To understand and interpret approaches to school evaluation as 'social constructs', this study is located within the social constructionist research paradigm (Burr, 2015; Holstein and Gubrium, 2008). As Burr (2015) explains, knowledge is constructed between people through their daily interactions, and it is in the course of these interactions that particular approaches to school evaluation may become dominant and official in the sense of being promoted by government officials, civil servants and other policy actors.

A social constructionist position enabled me to explore the perspectives of individual participants who engage in school evaluation as 'advisers', 'designers', 'implementers',

'researchers' or 'commentators' (see also Chapter 5). The experts who participated in my research provided me with their assessment of the evolving approach to school evaluation, based on their own 'constructs' and views on how schools should be evaluated. Social constructionism is a research paradigm which, in my understanding, encourages researchers to be cautious about their own assumptions about the social world (Burr, 2003: 3; Hosking, 2008; Pfohl, 2008). It also encourages researchers to take a critical position in relation to the social world. For instance, Hacking (1999: 6) explains that social constructionism is 'critical of the status quo'. To explain a critical social constructionist stance, Hacking explains that 'X' (a social phenomenon such as school evaluation) needs to be understood as socially constructed rather than inevitable. In this account, a particular approach to school evaluation (as 'X'):

> X need not have existed or need not be at all as it is. X, or X as it is at present, is not determined by the nature of things; it is not inevitable. (Hacking 1999: 6)

Based on Hacking's explanation, a particular approach to school evaluation is not inevitable due to some essential nature of education. Approaches to school evaluation have been shaped by social events, history, values, ideas and ideologies of powerful social actors. And yet, official approaches to school evaluation (i.e., the four *pillars* of school evaluation) may be taken for granted rather than considered to be socially constructed. However, just because a powerful policy actor promotes or even imposes a particular official approach, for a social constructionist researcher this does not mean that we should assume that there is no alternative way of evaluating schools. Hacking goes further by stating that a critical social constructionist stance also encourages a position whereby a researcher considers a possibility that:

> We would be much better off if X were done away with, or at least radically transformed. (Hacking 1999: 6)

Being open to this possibility enabled me to be open to a radical stance on the official approach to school evaluation adopted by some of the experts who participated in my study (see Chapter 7, Section 7.2).

When it comes to the ontological perspective of social constructionist researchers, two tendencies have been identified: 'strong' and 'weak' social constructionism (Burr, 2015; Loseke and Kusenbach, 2008: 514). Strong constructionism challenges all objectivist assumptions about reality and posits that everything is socially constructed (Burr, 2015). The proponents of weak social constructionism argue that there are pre-existing material realities about which agreement can be reached (Loseke and Kusenbach, 2008). I consider myself to be a 'weak' social constructionist researcher, as I believe that diverse views on social reality depend on a

range of factors such as level of expertise, gender, age, and professional position, as well as certain *foundational ideas* which Christie and Alkin (2013) have written about (see Chapter 3). In this, I agree with Burr (2015) that, as educators, we construct our own ideas about how schools should be evaluated.

It is important to explain that social constructionism is concerned with how knowledge is constructed and understood and, consequently, it is an epistemological rather than an ontological perspective (Andrews, 2012). Therefore, one of the aims of my research and the related research question 3 focused on evaluating educational experts' views on the official approaches to school evaluation in England and their suggested alternatives by considering these in the light of knowledge in the field of evaluation.

## 4.3 Expert interview

In this study, I employed the research method of expert (elite) interview to capture comprehensive perspectives and experiences of the participating experts who have worked in a range of roles in the English education system, as *policy actors*, *policy influencers*, and *school practitioners* as *policy implementers* (for details see Chapter 5). There are overlaps and differences between expert and elite interviews, but it is important to make a clear distinction between the two in order to demonstrate my reasons for choosing expert interviews as a research method. The elite interview derives from the field of sociology and unlike expert interviews, it has a 'long-standing tradition' as 'a basic form of qualitative interviewing' (Bogner et al., 2018: 2). An early example, which aimed to generate qualitative data, dates back to the 1970s and the work of Dexter, a political scientist and author of one of the first books about elite interviews. According to Bogner et al. (2018:2), Dexter's elite research:

> was driven by the idea that, since the power elite controls society to a great extent, empirical insights into the worldviews and interests of the elite are necessary to understand societal order and change.

The reason for the rise of the elite interview, according to Harvey (2010: 2), is related to the 'critical' role of 'elite workers' in shaping the policies and characteristics of organisations. Harvey argued that elite interviews are effective in generating 'novel' and 'insightful' data which may be difficult to obtain through other, more traditional interviewing methods.

The expert interviews are, by contrast, 'rarely thought-through and to a lesser extent methodologically reflected' (Meuser and Nagel, 2009, cited in Bogner et al., 2018: 2). Expert interviews are generally conducted to generate knowledge in the 'exploratory' phase of a

project, in order to triangulate quantitative research data or to 'structure the area under investigation and to generate hypotheses' (Bogner et al., 2018: 3-10).

In educational research, expert and elite interviews are relatively rare. For example, elite and expert interviews do not feature much in research methodology textbooks on educational research such as Cohen et al. (2007). The search on the ERIC database for 'expert interviews' featured just 12 peer reviewed empirical research publications in education which employed this method of data collection. Of those, only one focused in detail on the expert interview methodology (Döringer, 2021). Below, I discuss definitions of expert interview, outline how it may be seen to differ from an elite interview, and then justify why this method fits well in my research.

### 4.3.1 Defining 'expert interview'

The term 'elite' can mean many things in different contexts and holding an elite status may be embedded within geographical location and time. There is no clear-cut definition of the term 'elite' in published literature (Harvey, 2010; Richards, 1996). The existing definitions refer to conditions such as power, status, position, or knowledge. For instance, Dexter (2006: 19) refers to people who are 'the influential, the prominent, the well-informed'. Zuckerman (1972: 160) points out that holding a significant amount of power within a group is the condition to be an elite. McDowell (1998: 2135) describes elites as individuals who are 'highly skilled, professionally competent, and class-specific'. Woods (1998, cited in Harvey, 2010: 195) refers to elites as individuals 'who occupy positions at the top of the employment and income pyramid'. Welch et al. (2002: 613) describe elites as 'informants who occupy a senior or middle management position'.

Given this broad understanding across the social sciences, I decided to define the elite participants who participated in this study as experts. In the sources I have read on this type of interview, most authors refer to elite rather than expert interviews. My choice to refer to experts rather than elite participants follows Van Audenhove's (2017) argument that the reference to expert interviews shows the researcher's focus on the knowledge rather than power and status of such participants (see Figure 4.1 below). Although status and position were not irrelevant to the experts who agreed to participate in my study, the core aspect of 'expertise' in this study is expert knowledge. The sources of types of knowledge the participants possessed included historical, technical, theoretical, and/or practical knowledge gained through many years of working in different roles and at different levels of the education system. Therefore, I did not

rely purely on status and power, and even position, but on knowledge and experience. I would argue that being an expert in a field, gained through knowledge, remains even if the elite or expert status is diminished or undermined. This has been the case with the expert status of teachers and school leaders in the policy environment post Education Reform Act (ERA) 1988 (Gunter and Forrester, 2009).
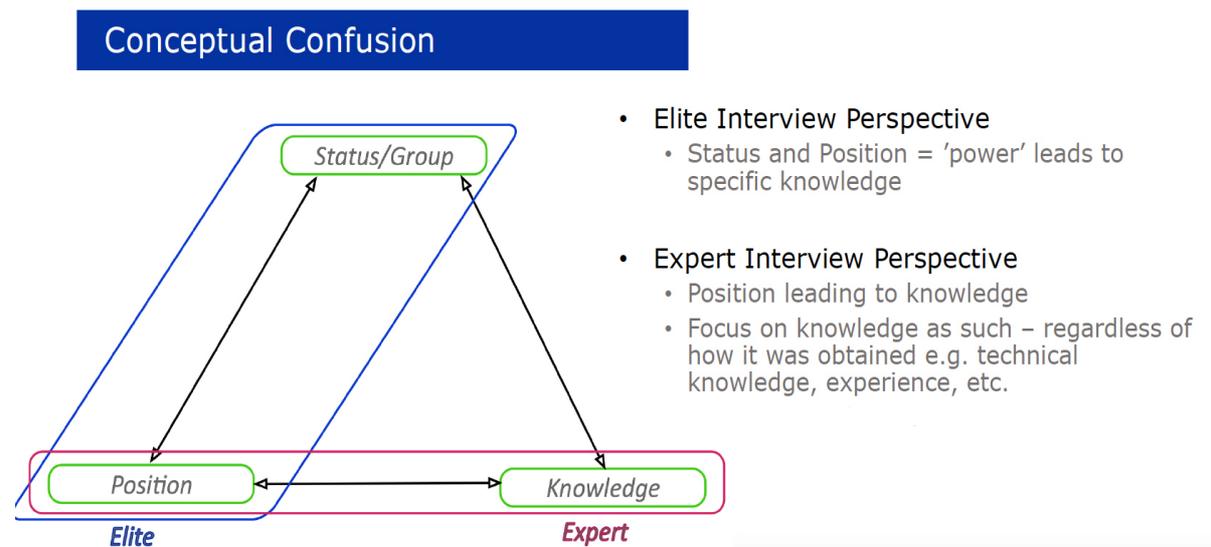


**Figure 4.1:** Van Audenhove's (2017) conception on expert and elite interview

Although expert interview was deemed to be an appropriate tool to address my research questions, I am aware that expert participants provided me with subjective accounts of their experience and views on approaches to school evaluation. However, I did not conduct this study with a view to establishing 'the truth' in a positivist manner (Richards, 1996: 200) as explained in Section 4.2. As a social constructionist researcher, I have not sought to discover 'objective' knowledge or 'neutral' facts. Instead, I have been interested in expert knowledge because expert knowledge 'determines social practices and institutions to a certain extent' (Bogner and Menz, 2009, cited in Bogner et al., 2018: 5). In other words, experts influence, make and/or enact approaches to school evaluation and their knowledge and experience can shed light on their benefits, limitations, and more viable alternatives.

Although some argue that "experts are not primarily interviewed because of an interesting 'solid' or canonical knowledge as one can find in handbooks and encyclopaedias" (Bogner et al., 2018: 5), what can be learnt from an expert is related to questions which have been asked.

As Richards (1996) suggested, experts may offer their expertise in interpreting documents or reports, explaining the outcomes of policies, as well as provide information not recorded elsewhere, or not yet available to the public. Thereby, experts in my research enabled me to answer my research questions by appraising the changing approaches to school evaluation in England of the last 30 years, offering me depth and insight, from a number of perspectives related to their key roles in the education system. Their expertise has been developed through research, at the level of policy and/or through working in schools, in the roles of *policy actors, policy influencers* and *school practitioners*. Chapter 5 elaborates on these roles in more detail.

### 4.3.2 Expert interview as sensitive research

Expert interviews can be seen as 'sensitive research' (Cohen et al., 2007: 127), i.e., research which is 'conducted on, or with powerful people, those in key positions, or elite institutions.' As Cohen et al. (2007: 128) explain, policy-related research is sensitive because it deals with issues 'about which there is high-profile debate and contestation, as issues of a politically sensitive nature'. Similarly, interviewing experts whose professional roles rest on making, influencing, or enacting policy made my research sensitive. The sensitivity of my research also stems from interviewing participants who are public figures and key decision-makers at the level of the education system and are, therefore, accountable at a high level, to the government and the public. For instance, one of my participants made an important decision that influenced the educational sector 'tremendously', a decision that was misunderstood. This participant also explained their regret about making this decision. To protect the confidentiality of this participant, I am unable to provide more detail.

Researchers have to be 'acutely aware of a variety of delicate issues' which might stem from the nature of sensitive research (Cohen et al., 2007: 119). This suggestion resonates with arguments about expert (elite) interviews. For instance, Harvey (2010, 2011) and Richards (1996) argue, based on their personal experience of conducting elite interviews, that researching with elites (experts) can be challenging and might require some specific additional strategies. Importantly, data collection through both elite and expert interviews faces 'similar methodological challenges' (Litting, 2009, cited in Bogner et al., 2018: 2), which I followed in this study. Although Dexter (1964: 557, cited in Harvey, 2010: 5-6) argued that junior researchers should avoid interviewing elite members because they are 'ill-prepared' and 'needlessly take up the time of important persons', I made sure that I very carefully prepared for each of my interviews.

## 4.4 Planning and conducting expert interviews: challenges and ethical considerations

This section reflects on my experience of interviewing expert participants who are often top authorities in their role and field of expertise and the challenges which I faced, not only due to the nature of expert interview but also due to who I am - a female junior researcher with an international background. Where published literature sources refer to 'elite' rather than 'expert' interviews, I follow the authors' original terminology.

### 4.4.1 Planning expert interviews

The planning step in sensitive research requires researchers to 'demonstrate a great deal of ingenuity and forethought' (Cohen et al., 2007: 122-123). The first step in planning my interviews involved identifying a list of experts and creating a corpus of their work for analysis. Therefore, my first task was to establish the critical or key informants with whom initial contact would be made, because such contacts would put me in touch with more contacts. At this stage, I followed the media, read the literature, and asked my supervisors' advice. I then identified a number of key authorities in school inspection and education policymaking. Those experts had authority in their specialism with their comprehensive knowledge of educational evaluation and assessment policies since 1988, both internationally and nationally. I did not limit my initial choices and their current or former role and I was interested in a wide range of roles that enable professionals to gain expertise: academics, school practitioners (i.e., Senior Management Team members and senior school leaders), members of professional associations (i.e., National Association of Headteachers, The Association of School and College Leaders), policy actors (i.e., Regional School Commissioners, Ofsted Inspectors) and researchers from think tanks.

I then listed and reviewed their work, if available (e.g., publications or public statements), in order to consider who might be approached to participate in my study, as the most knowledgeable participants with expertise in school evaluation and in-depth knowledge by virtue of their professional role, authority, access to networks or experience. At the stage of shortlisting, I also tried to identify diverse groups of participants, representing diverse career trajectories. At this stage, some participants' email addresses were not readily available. As part of my purposive sampling (Cohen et al., 2007), I also tried to identify experts who were supportive of the official approach to school evaluation, as well as experts who appeared to be critical of this approach. This was in most instances through searching online for their

'presence' and/or record of publications. Specifically, in the case of experts active on social media or whose profiles and statements were available publicly on website pages or e-newspapers, I could ascertain their commitments to different forms of evaluation on the basis of their tweets, speeches on YouTube, podcasts, blogs, and public campaigns. In the case of academic experts, I was able to gain insight into their commitment on the basis of their publications. For example, upon reading Orion's publications, it was apparent that he was critical of evaluation approaches introduced under the influence of New Public Management in England. In the case of participants who did not have a social media presence or a record of publication, I did not know their commitment prior to the interview but I sensed their views on the basis of what was told about them by the people who referred me to that particular participant. For instance, Ned was suggested to me by Nora. Nora already knew about Ned's school and his reputation for 'progressive and holistic' student assessment practices at his school - by involving not only teachers but also parents, the students themselves, and their peers. This information enabled me to anticipate that Ned's views could be more progressive. Indeed, he was the only participant who suggested an evaluation system without Ofsted.

At this stage, I also designed a semi-structured interview schedule (see Appendix 3). First, I prepared a list of interviews questions, avoiding closed questions. I used those questions to shape the overall interview structure, but I also tried to follow the 'flow' of the conversation and asked additional questions or prompt questions. This was to enhance the gathering of depth information from the participants' specific specialism. Both the structure and flexibility of this type of interview made it possible to gather comprehensive data, allowing me as the researcher to focus on essential points guided by prompts and follow-up questions to participants, as well as major questions (Thomas, 2013).

### 4.4.2 Ethical considerations

The consideration of ethical issues was of high importance at every stage of my research, from gaining ethical approval from the Research Ethics Committee to the data collection process, because of the nature of the expert interview discussed above. To gain the approval of the Ethics Committee of the School of Education and Lifelong Learning at the University of East Anglia (UEA), I rigorously followed the ethics guidelines set by the UEA (2013) and the British Educational Association (BERA, 2018). These guidelines included informed consent from participants through a consent form which detailed the aims of this research, the proposed use of the data and the right to withdraw from the research. Because of the potential of some

of the participants being identifiable due to the uniqueness of their position or achievement, I prepared two different consent forms (see Appendix 1). One form was for experts who might be identifiable, despite all my efforts to anonymise their identity and one for participants who would not be identifiable. As can be seen in Appendix 1b, the form for participants who might be identifiable refers to this possibility (and I also highlighted this at the beginning of the interview). Although the potentially identifiable participants have given me consent to use their data in this thesis and in future publications, I deployed additional strategies to further protect their anonymity. My strategies to protect them varied from redacting some of the information they shared to omission of identifiable information such as the specific role. Therefore, some details and 'potted history' of some participants presented in Chapter 5 have been redacted to protect the anonymity of all participants.

### 4.4.3 Conducting expert interviews

As noted above, I chose my participants purposively to build up a sample that is satisfactory to the specific needs of my research (Cohen et al., 2007). I also benefited from a snowballing technique to access more experts but used it with caution by scrutinising online whether the expert who was recommended by a particular participant would offer the depth of expertise and a range of roles to ensure a comprehensive data set. The size of the sample is determined by saturation which is the point at which obtaining new information is unlikely, and there is no requirement for replication (Cohen et al., 2007). At the point when 15 interviews were completed and transcribed, I interviewed for the total of 21 hours and 18 minutes, and the total wordage of interview transcripts reached 140,356 words. For Harvey (2010), the number of interviews is not necessarily an indicator of high-quality research because of issue is also the length of interviews and the insight offered by experts. The table below shows the total number of words for each interview transcript, the length of each interview, as well as the range of professional roles of the participants. The status of member checking and redaction requests can be found in Appendix 3.

| Category of expertise | Name | Time (21h 18m) | Words (140,356) | Date | Area of expertise |
|---|---|---|---|---|---|
| Policy Actor | Richard | 01:16:15 | 7,175 | 09/11/18 | Inspector, policy maker- international and the UK |
| | Cooper | 02:03:04 | 11,239 | 12/12/18 | LA management roles, academic |
| | Dennis | 00:57:40 | 6,379 | 18/12/18 | HMI, policy adviser, researcher |
| | Gabriella | 00:46:18 | 6,080 | 26/02/19 | Inspector from a devolved system |
| | Bella | 01:29:49 | 11,162 | 07/06/19 | HMI, adjudicator, policy adviser |
| | Neil | 01:26:00 | 9,502 | 26/06/19 | Policy adviser and policy maker |
| Policy Influencer | Nora | 01:19:12 | 11,051 | 25/10/19 | Researcher and academic in the field of assessment |
| | Gordon | 01:18:44 | 9,517 | 18/01/19 | Researcher, specialised in Ofsted inspection, and its alternatives |
| | Eduardo | 01:10:58 | 9,279 | 14/06/19 | Researcher, policy adviser |
| | Orion | 02:47:49 | 12,682 | 16/10/19 | Researcher and academic in the field of evaluation |
| School Practitioner | Kelvin | 01:49:51 | 13,136 | 10/10/18 | Experienced headteacher |
| | Torr | 01:09:23 | 6,926 | 13/03/19 | Senior teacher with academic background |
| | Felicia | 01:04:34 | 8,771 | 28/03/19 | Deputy head- international and the UK |
| | Ned | 01:07:07 | 7,332 | 5/03/19 | Headteacher, lead headteacher, national leader of education |
| | Kent | 01:32:17 | 10,125 | 17/05/19 | Principal, Ofsted Inspector |

**Table 4.1:** Interview time and word count

Access is a significant concern in research with experts and it depends a great deal on 'serendipity' and 'social networks' (McDowell, 1998: 2135). Access could be even harder if the issues being researched are controversial and include sensitivities around power. For instance, Cohen et al., (2007: 127) state that policy-related research with 'powerful people' is sensitive and is 'frequently refused' as those people 'may not wish to disclose information' which may hurt them. Controversies around official approaches to school evaluation in England and participants' security needs may explain why my invitation was rejected by potential participants on a number of occasions. Chamberlain and Hodgetts (2018: 15) wrote about collecting qualitative data with 'hard-to-reach groups' and stressed the importance of enhanced ethical consideration:

> approaching data collection with hard-to-reach groups involves actioning an ethics of reciprocity that facilitates a forging of closer human relationships with participants and others whose decisions impact on the lives of these groups and brings benefits to all parties. In working this way, barriers to working with hard-to-reach groups tend to dissipate.

To build rapport and trust, I tried to be 'transparent' (Harvey, 2011: 5) and carefully crafted my initial email communications to explain who I am, where I was studying, under whose

supervision, what the nature of my research was, who was sponsoring me, as well as how long my interview would take, how the data would be used, how the results would be disseminated and whether the information would be attributed or anonymous (Harvey, 2010: 19; Harvey, 2011: 5). Preparing personalised invitation emails which outlined my knowledge of participants, including their projects, campaigns and/or published work helped me to establish rapport, in addition to gaining access. An example of these personalised emails can be found in Appendix 3.

I also gained access to participants through introductions by others, for example, a person who was involved in national policy-level research introduced my research to a senior-level person who worked for Ofsted. This contact then supported me with accessing more contacts at Ofsted. I was even asked in which specialism I need a person who works for Ofsted to contribute to my research. My other strategy was to introduce myself and my research to participants at various education events. For instance, I attended a summer school at University College London which was for inspectors from all over the world, i.e., India, Pakistani, Turkey, England, Scotland, New Zealand, as well as various conferences and seminars (i.e., in Hamburg, Amsterdam) and met some of my participants at these events. Although these strategies might question the representatives of my sample, they enabled me to collect rich, depth data. Importantly, as suggested by Harvey (2010: 8), pursuing multiple avenues for gaining access to elite populations may 'reduce the potential bias of only speaking to people within a particular social network'. Whilst the above strategies were helpful, I am aware that the participants who were able and willing to participate in my research worked in the roles of *policy actors*, *policy influencers*, and *school practitioners*. Despite several attempts, I could not establish contact with experts employed by professional associations. This might be related to the fact that the 'voice' of professional associations is already publicly and widely available through their reports, similar to publications by academics and media coverage of policies. However, the more personal voice of policymakers and academics which I am able to report in this thesis is scarce in the public domain, as they represent their affiliated organisations and professions rather than their personal views. My easier access to *policy actors* compared to leaders of professional associations might be also related to the snowballing technique I used. According to Cohen et al., (2007: 116):

> maybe because it is a sensitive topic (e.g., teenage solvent abusers) or where communication networks are undeveloped (e.g., where a researcher wishes to interview stand-in 'supply' teachers – teachers who are brought in on an ad-hoc basis to cover for absent regular members of a school's teaching staff – but finds it difficult to acquire a list of these stand-in teachers), or where an outside researcher has difficulty in gaining

access to schools (going through informal networks of friends/acquaintance and their friends and acquaintances and so on rather than through formal channels).

### 4.4.4 Challenges

I conducted interviews both in-person ('face to face'), by audio-calling (i.e., telephone, online audio-calling/camera off) and by video-calling (i.e., Skype) depending on the participant's preference. I have found the in-person and video-calling interviews to be more effective in bringing more depth information than audio-calling interviews and this may have been the result of the rapport between myself and the participant. My experience in this regard was in line with the literature; for example, Sturges and Hanrahan (2004) explain that telephone interviews restrict non-verbal elements like eye contact, which is a disadvantage for effective communication. Further, compared to face-to-face interviews, telephone interviews allow for less control of the interview situation. In my experience, the telephone interviews were shorter than the face-to-face ones, compared to in-person and video-calling interviews. However, I found the video calling interviews as effective as the in-person ones, because there was still a virtual visual presence and eye contact. I also benefited from the video-calling online interview as it was easier to set a time with busy experts and a cheaper option for me as a researcher based in one part of the UK and yet able to interview participants in many different parts of the country.

Whilst online video-calling interviews worked for me, whether interviews are in-person or virtual can affect the interview process and the data collected. For instance, some researchers find online interviews challenging because of potential technological malfunctions or participants' lack of expertise with the technology (i.e., Mirick and Wladkowski, 2019). However, I did not experience technological malfunctions such as internet disconnection. I found the effectiveness of online video calling interviews similar to in-person interviews. However, this might be related to offering participants a choice about their preference how to meet. What is more, I have also found that online interviews were more useful. For instance, one in-person interview was in a café, because this was preferred by the participant, and this caused low quality recording due to background noise. I could not transcribe her interview and could not report it in this thesis as data. I believe that, if I had a chance to do this interview in a virtual form, I would have used an in-built recording function of an online tool (i.e., MS Teams) and this would have given me a better-quality recording. Importantly, some studies suggest that although in-person interviews have been claimed as the 'gold standard' for social science research interviews before the pandemic, the 'widespread use of online-based

interviewing methods will likely endure as equivalent to in-person methods' (Owens, 2022: 121). So, online interviews are not something we can avoid, or we should try to avoid, but instead online interviews are something we need to get used to it.

I gained permission from every participant to record the interviews for later transcription. I was never refused to record or provided information 'off the record' (Cohen et al., 2007; Harvey, 2011). This enabled me to create a verbatim transcript of each interview. As part of my preparation for interviews, I tried to respond with sensitivity to participants 'hedging' their answers and respect this. Whereas pilot interviews are recommended (Cohen et al., 2007: 129) as beneficial for conventional interviews, there is very little guidance concerning whether interviewers should conduct a pilot work with elite or expert participants (Harvey, 2010). I did a mock interview with my friend, and subsequently I reorganised the wording and order of some questions. However, the actual interviews with experts showed me another side of the coin. Some expert participants exercised control over which questions they answered and the terminology they used. For instance, when asked about 'high stakes testing', one of my participants challenged me and said, 'this is nonsense' and 'why you think it is high stake testing?' She then challenged my reference to the 'accountability system' asking me: 'what do you mean by accountability?' and 'I have not heard that marketisation is used in educational settings'. Although high stakes testing is a notion widely used in published research and literature, I did not anticipate that my use of this, and other terms, might prove to be contentious to some participants. In that interview, I responded by explaining that this notion does not belong to me and provided her with some examples from the literature. This experience taught me the importance of responding to, and preparing for, criticism. I revised my questions to use a more neutral terminology and tried to refrain from using potentially controversial terminology until further in the interview. During some interviews, I rephrased my questions to reflect the participants' use of terminology. For instance, once the participant used the notion of league tables or performance tables, I kept using such terms accordingly. I also expressed contentious notions by referring to the: 'so-called… by some'. For some terms that cannot be expressed neutrally, I added citations to my interview questions and emailed such questions prior to the interview. This was meant to deflect participants' attention from personal sensitivities. A similar experience was shared by Harvey (2011: 12), during his interview with a CEO who became 'extremely agitated' saying that his questions were 'so vague' and 'not relevant', and his responses to his open-ended questions were extremely short. This experience even 'hampered' his confidence. Harvey suggested that researchers new to interviewing should

be conscious of how to cope with difficult interviews.

The literature also reports that characteristics of the researcher may affect the interviewee, i.e., 'sex, race, age, status, clothing, appearance, rapport, background, expertise, institutional affiliation, political affiliation, type of employment or vocation' (Cohen et al., 2007: 130). As for gender, the literature on interviewing powerful people suggests that females are in a favourable position in both access and during the interview. Contrary to the literature, I did not notice that I was treated either favourably or unfavourably because of my gender. Walford (1994) reports that female researchers may be at an advantage in that they are viewed as more harmless and non-threatening.

The literature also points to the appropriate length of an interview, though does not indicate what this length could be in the case of elite interviews (Harvey 2010, 2011). On average, my interviews lasted 1 hour and 25 minutes. The experience of Harvey (2010) was to some extent similar to mine; Harvey asked for about thirty minutes, but the interviews typically lasted for approximately forty-five minutes. Interestingly, many of my participants expressed their wish to talk for longer than forty-five minutes or an hour. After I reminded them that an hour had passed, almost all of my participants asked me to continue and confirmed that they are willing to talk longer. This could be explained in many ways such as my characteristics (i.e., my gender, being a junior researcher) or by the fact that the experts had more time on the day. However, I also believe that inviting them to participate in my study as expert witnesses and assessors (of the 30-year history of school evaluation England) seemed to offer them an 'opportunity' to share their views in a more personal way. Some participants appeared to be waiting to 'be asked' further questions, particularly about alternative models and approaches to school evaluation. An example exchange with one of my participants (Cooper) illustrates this phenomenon. Cooper, who had been employed in a range of important roles since the 1980s, told me: 'I spent my whole life in this area. So, I want to talk.'

Preparing questions and finishing the interview on time are important, however, equally important is being able to turn the interview into a 'conversation' or 'dialogue'. According to Bogner et al. (2018: 12), taking time to listen to the interviewees is also important, even if they might digress and mention aspects of no immediate relevance to the central research topic, especially if the researchers are interested in the rich investigation of 'interpretative knowledge'. This is why, after a couple of interviews, I noticed how it was more appropriate to be open to the flow of the conversation and show my willingness to hear more. On the other hand, because

of time limitation, it was also important to move the conversation to the next question by briefly reflecting on the interviewee's ideas. This is because some participants were very keen to talk about their long-term experience.

Research that employs expert interviews entails encounters with some 'powerful people' or people of authority, thus involving power relations between the researcher and the interviewee (Elmwood and Martin, 2000: 650). As suggested by Cohen et al. (2007), where the interviewer has a lesser status than the respondent, 'managing power relations' is very important (p.123) because, when respondents feel that they are interacting with a 'lowly novice' researcher, they may use the interview as 'an opportunity for lengthy and perhaps irrelevant self-indulgence' (p.131). In response to this insight, I tried to establish my own status by explaining, in every email invitation, why I am researching this topic, which is to contribute Turkish education system as a prospective policymaker. I hoped that presenting myself in this way could reduce the power gap and highlight that my research seeks to contribute to a country that requires expert knowledge in order to develop a better evaluation system. I think stressing the benefits of research is very important in an expert interview because most experts have a very tight schedule, and they may want to make their effort worth it. In fact, without prompting, some of my participants provided suggestions from good practices of England, for instance:

> Can I throw one thing in? [I said yes please]. Which is, I think, the secret or accidental benefit that Ofsted has is that most of our inspectors are serving practitioners in schools. So, sixteen hundred inspectors are heads… that's their day job. They only inspect for us nine days a year. It's not a very economic model but they all get all our training. And so, it's establishing a consensus around certain education ideas. And so, if you're thinking about culture change, there's something about everyone having been through the same conversation and training. And that creates the fertile ground for then moving the system forward. That's my thought on that. (Neil)

To prepare for power differential, I also completed meticulous preparation for every interview, by 'doing my homework': engaging with the participants' publications, work, and campaigns, so that I went into each interview with some knowledge of each participant or some knowledge of the demands of his or her role. I also believe that referring to their work in my questions helped to build rapport. As noted by Harvey (2011: 6), the researcher should be ready to be asked questions and use this as an opportunity to 'project a positive impression in order to gain their respect'. In a similar vein, Bogner et al. (2018: 12) suggested that 'the interviewer should deliberately demonstrate his or her own expertise in order to gain the recognition of the interviewee'.

I must admit, however, that I noticed this after the first two interviews. According to the culture

in which I was raised, it is important to be humble and show explicit respect to the 'older generation' or to the people who have a 'higher position' by acknowledging my junior researcher position. But this sometimes did not work in this type of interview. In my early interviews, my stance was perhaps one of 'excessive' respect which, in turn, may have undermined my own knowledge. For instance, when interviewing a highly esteemed academic, whose works greatly influenced the international, national, and local level education policies for over 50 years, I felt challenged when I was asked questions such as 'whether I read book X or article Y' or whether 'I know person X'. I felt I was put in the spotlight and my knowledge was 'tested'. This made me aware that 'doing my homework' may not be enough and gaining reassurance and encouragement from my expert participants would help me, as well as other novice researchers in my position. This reassurance was offered to me by another interviewee who, before sharing his views, asked me a question about Victorian times in education. I told her/him that I am aware of the history of education in England since the 1960s and I assured her/him that I would deepen my knowledge about the educational history of England before the 1960s. The interview proceeded and I was not presented with further suggestions pointing out 'gaps' in my knowledge. As noted by Abels and Behrens (2009, cited in Bogner et al., 2018: 11-12), preparing for expert interviews extends to being aware that:

> In some cases, he or she [expert] even displays a patronising attitude towards the interviewer, attempts to show how well disposed he or she is, and to dictate the content of the conversation to the (seemingly) inexperienced or inferior interviewer – often with a gender-specific bias if a young female researcher interviews an older male expert or member of the elite.

On the other hand, 'an asymmetrical interaction situation where the interviewer is seen as inferior or naïve is not generally problematic… naïve questions stand a good chance of producing the most interesting and productive answers' (Bogner et al., 2018: 12). Therefore, in many instances, the data I gathered became more productive: 'a naïve interviewer is seen as especially trustworthy' (Abels and Behrens, 2009, cited in Bogner et al., 2018: 12). On the other hand, the disadvantages of interviewing experts, as noted by Bogner et al. (2018: 12) may arise when interviewees present researchers with:

> interminable monologues about trivia or things they already know, they plod through the contents of textbooks, or retreat to common places. There is hardly any likelihood that difficult specialist issues can be clarified since it is easier to ignore supplementary questions.

Therefore, there is no 'best practice' concerning the interaction in interviews with experts and responses by both parties depend on the situation. This requires skills that, as a junior

researcher, I started developing after the first few interviews.

The chosen location may affect many aspects of an expert interview. For instance, Harvey (2010) argued that the answers of participants such as what s/he wants to disclose are influenced by location. In agreement with Cohen et al. (2007: 128) I have experienced that research with powerful people usually 'takes place on their territory, under their conditions and agendas,' and that would be 'disconcerting for researchers'. What is more, as claimed by some researchers (e.g., Elwood and Martin, 2000), the choice of location would affect the researcher's success in gaining access. In that sense, it would be best to leave this decision to the expert, as I did, but perhaps note the negative consequence of researching 'on their territory'. This, however, may limit the number of experts to access. In one of my interviews which I conducted at a cafe, the recording quality became extremely poor, because of background noise. While some authors (Elwood and Martin, 2000; Harvey, 2010) argue that interviews outside the workplace are easier to expand as well as to broach more confidential information, Dexter (2006: 48) suggests prioritising the workplace because some interviewees 'will tolerate interruptions [it comes from for instance family members] which they would not in their offices'. In this regard, I would argue that this is a decision that needs to be made together with the participants if possible, considering the strengths and weaknesses of the decided location.

Although the reason for interviewing experts was to explore their conceptualisations of school evaluation, sometimes the conversation may turn out to be about 'convincing the researcher' rather than 'sharing views with the researcher'. This might be the case because the topics around school evaluation are inherently political, and policymaking and implementation require convincing others. I noticed, during some interviews, that challenging the participants' views was not productive and I felt that at times the participants were waiting for an answer or affirmative reaction to their arguments from me. Some participants even stopped and asked me directly what my idea was about their argument. In these situations, I sought to respond tactfully by repeating their points and prompt elaboration on their part. This facilitated a more open conversation between us rather than giving closed answers and as suggested by Whitty and Edwards (1994: 22, cited in Cohen et al., 2007: 122). I also refrained from declaring my own views, when asked, or referred to general answers, such as 'some thinking is different… What would be your views about these alternative arguments?' My position was based on a conscious effort to be 'neutral' and open to diverse positions taken by my research participants,

so as to learn from diverse views.

Another point to consider is that 'some respondents may be unwilling to disclose sensitive information, particularly if it could harm themselves or others' (Cohen et al., 2007: 333). In other words, their answers would be constrained by their professional confidentiality. As was noted by Fitz and Halpin (1994: 40, cited in Cohen et al., 2007: 128), they 'glimpsed an unfamiliar world that was only ever partially revealed', and one in which they did not always feel comfortable. One of my participants refused to answer my question about marketisation as this would show his/her political views and that was not appropriate for their current role. I accepted their answer and moved the interview on:

> Researcher: If you had a magic wand in your hand, would you like to leave behind these marketisation-based strategies?
>
> Respondent: That's not a question I am going to answer. Because that's a political question and I am a civil servant, and I cannot answer political questions.

Also, cross-cultural interviewing issues such as language differences and different norms and values between my participants and me (Patton, 2002: 391-393) existed. This is however generally a neglected problem in qualitative research and expert interview is not exempt from this (Bogner et al., 2018: 13). For instance, I experienced some challenges pertaining to the use of many idioms during interviews. I overcame this issue by honestly asking participants for further explanations regarding their argument and this very much improved the quality of the ensuing conversation.

## 4.5 Data analysis and reporting of expert interview

There is no standard procedure for analysing expert interviews or a correct way to do it and present the findings (Bogner et al., 2018: 16) but data analysis 'should abide by the issue of fitness for purpose' (Cohen et al., 2007: 461). I employed thematic analysis (Braun and Clarke, 2006: 37) because this method 'can usefully summarise key features of a large body of data, and/or offer a thick description of the data set' and 'highlight similarities and differences across the data set'. The focus of thematic analysis, as explained by Namey et al. (2008: 138), is concerned with:

> identifying and describing both implicit and explicit ideas. Codes developed for ideas or themes are then applied or linked to raw data as summary markers for later analysis, which may include comparing the relative frequencies of themes or topics within a data set, looking for code co-occurrence, or graphically displaying code relationships.

I started the data analysis process by familiarising myself with my data. I first transcribed the

recorded interviews verbatim without reference to any nonverbal communication. Then I emailed the interview transcripts to each interviewee to ask their approval for the accuracy of the transcripts, along with a reminder about their right to withdraw any of the data if they wished. While some participants chose to have some text removed, most participants approved the transcript as transcribed (see Appendix 3). Only the redacted interviews have been used as part of this research.

I then familiarised myself with the data through a repeated reading of transcripts, writing a summary of each interview and writing analytic memos (see Appendix 4 for a sample excerpt of an interview transcript with initial codes and memos). Also, listening to the recordings enabled me to write further analytic memos. The recordings helped me to remember my initial thoughts about the experts and their views. Although I tried to complete the transcription as soon as possible after each interview, this was sometimes not possible. However, the initial thoughts and reflections should not be neglected and listening to the recordings allowed me to remember these initial thoughts. Overall, immersing myself in my data made my data coding and thematising smoother as 'searching for meanings, patterns and so on' started with data familiarisation (Braun and Clarke, 2006: 16). The following is an example of the analytic memos in response to the interview question (3) *What are your views about Ofsted?* (Table 4.2)

I then started organising the data and generating initial codes with an inductive, data-driven approach and used NVivo-12. The NVivo-12 software helped me to give equal attention to each data set and helped me with the coding process. Some codes turned out unique to an individual participant, but some were common to all, or most, of the interviews. However, I gave equal importance to each code because my aim was not to quantify the frequency of the arguments but to explore the comprehensive views expressed by the participants. For instance, when I was immersing myself in the data, reading and re-reading interview transcripts to identify initial codes and note patterns of similarity and difference across participants' answers, I tried to focus on the data rather than the conceptual framework developed in Chapter 3 of this thesis. In fact, at this initial stage of data analysis, my conceptual framework was yet to be developed. This allowed me not to be influenced by evaluation theories and helped me to code interview transcripts in an inductive, data-driven, way. When coding transcripts pertaining to Ofsted, I noticed a pattern in participants' responses. While some participants appeared to suggest that the negative consequences of Ofsted inspections are inevitable, others seemed to

be referring to those as a systemic problem. For instance, Neil noted that 'there are unintended consequences of whatever we [Ofsted] do'. This made me think that, according to Neil, the issues regarding inspection are not related to the inspection methodology nor to the purpose of inspection. In contrast, Felicia found inspections to be a 'morale-killing' experience, and I thought this was constructed as a systemic problem because of her reference to Ofsted's methodological weakness of comparing schools using a single framework as if every school has the same context. The initial codes are summarised in Table 4.2 below.

| Interview extracts pertaining to unintended consequences of Ofsted inspections | Analytic memos | Initial codes |
|---|---|---|
| There are unintended consequences of whatever we [Ofsted] do. (Neil) | unintended consequences are 'natural' because people do not like being inspected | Inevitable unintended consequences |
| You hear stories, I'm sure. Ofsted inspection teams or inspection is going to schools and being a horror story. You know, the terrible things they've said or done or whatever. I've never experienced that… All the inspection teams have been open and transparent and whatever… But those persons whose inspection experience has been positive, you're going to get a reasonable response… And the experience of being inspected is nerve wracking, it's frightening, it's all of those things. But you have to get over that a little bit… I think, we all big enough to cope with that. (Kent) | unintended consequences are an individual (rather than systemic) matter | Inevitable unintended consequences |
| I think, one of the upshots of a lot of developments after 1988 was a reduction in creativity. And also, the whole external pressure which was linked to people losing jobs. So, I was sitting in a meeting in this city with headteachers and then a couple of months later we had another meeting, and someone had disappeared and then another one had disappeared… headteachers were losing their jobs as a direct result of this and the pressures associated with it either through illness or because they were removed or because they got bad Ofsted and then they were removed, or they left before they have to. That's a sick system, it's not a healthy system... it is complete opposite of autonomy, mastery and purpose… (Kelvin) | unintended consequences are systemic ('that's a sick system') and stemming from the prescriptive Ofsted frameworks | Methodological problems cause the unintended consequences |
| I worked for a UTC in the last couple of years… Now, that's a specialist school, and intakes in year 10. 98 percent boys and focuses on engineering. That school got placed into 'requires improvement'… because it didn't fit the criteria that that school was never going to fit… It was morale-killing. But that's because they were using an evaluation process that compared us to schools nationally. And we were not a national school. Ofsted works on the basis that you have a nationally representative sample. What school does? Where is that school? (Felicia) | inspections as 'morale-killing' experience - a systemic problem because Ofsted compare schools but not every school has the same context | Methodological problems cause the unintended consequences |

**Table 4.2:** Example of the analytic memos in response to the interview question (3)

I then re-considered the initial codes to group them into higher-level patterns or themes. This involved sorting the different codes into potential themes and collating all the relevant coded data extracts within the identified themes. For instance, as can be seen in Table 4.3, codes such as the 'punitive enactment of accountability' and 'outcome-based accountability' could be combined under a higher-level sub-theme of: 'problems with accountability'. In a similar vein, I combined the codes pertaining to the essential nature and usefulness of accountability as other sub-themes. At this point, it became clear to me that 'accountability' needs to be a major theme to work on, as a complex and often contested notion. The data analysis process also consisted of reviewing the constructed themes in light of my research questions. For instance, though highly important, the purpose of education and political interference in education were two themes which I have identified in some of the interviews but did not have scope to discuss in much detail in this thesis. I was careful to create coherent but also diverse themes. The themes finally established included: standards, accountability and parental choice and I used them to guide me to analyse and interpret the data. Table 4.3 below illustrates the codes identified in the interview data under the theme of accountability.

| Theme | Sub-themes | Codes |
|---|---|---|
| Accountability | The essential nature of accountability | Problems with accountability before 1988 |
| | | National accountability to reduce difference between schools |
| | | External accountability as an effective driver of school improvement |
| | | Evolving accountability – changing for the better as it has evolved |
| | Problems with accountability | External 'design' of the official accountability system as a driver of school improvement |
| | | Outcomes-based accountability |
| | | Accountability to whom: accountability to Ofsted |
| | | The punitive enactment of accountability |

**Table 4.3:** NVivo codes under the theme 'accountability'

I defined and named themes by identifying the 'essence' of what each of my themes is about and which aspect of the data each theme captures. For each individual theme, I wrote a detailed analysis and identified the 'story' that each theme tells in relation to my research questions (Braun and Clarke, 2006: 22).

Although I sought to identify some patterns in the views of *policy actors, policy influencers* and *school practitioners* (see Chapter 5) depending on these roles, there seemed to be no pattern. However, I soon noticed a 'polarisation' in the views of participants who generally

advocated for the official approach since 1988 and those who took a predominantly critical stance in relation to the official approach. In other words, the data pertaining to their overall views on the official approach to school evaluation appeared to be either predominantly positive (Bella, Kent, Neil, Eduardo) or predominantly critical (Cooper, Ned, Torr, Dennis, Gordon, Kelvin, Nora, Felicia, Orion). For example, Kent's view was predominantly positive, both when he made the following point and in the remainder of the interview:

> I certainly think, before 1988, you know, before evaluation, there were some really bad schools, really bad. And I'm not saying there aren't still some really bad schools. But there are far, far fewer now than where there were. And I don't think there's anybody that could argue that actually that hasn't been in large part due to the inspection regime that we have. (Kent)

By referring to the time before 1988 as the time 'before evaluation', Kent appeared to hold the view that there was no evaluation prior to the Education Reform Act (ERA) of 1988. In contrast, Cooper appeared to be predominantly critical in his views on the official approach:

> what we have introduced in the name of accountability is inspection, performance tables and alongside that is parental choice… The accountability system is an external driver, and it isn't the best mechanism for improvement… after 30 years of doing things this way, it's increasingly going down the same road, it hasn't improved things. But, at some point, we'll realise it hasn't improved things… there is a saying: 'I don't fatten my pigs by weighing them.' Just measuring something doesn't change it. You can't make a pig fatter by weighing it… It's about improvement, not about judgements. Judgements don't improve. (Cooper)

However, for two of the participants (Richard, Gabriella), it was difficult to ascertain whether or not their views were predominantly positive or critical. For example, Gabriella was invited to participate in this study 'purposively', in order to share her expertise in alternatives and lessons from other countries (research question 2). Thus, the questions asked in the interview with Gabriella were tailor-made to her specific area of expertise rather than her views on the English approaches to school evaluation. Richard was generally supportive of the four *pillars* of school evaluation, though also stated that he preferred to stay neutral in relation to the current official approach to school evaluation, as he worked in England over 15 years ago. However, he presented some suggestions for how to improve the official from a pragmatic stance (Chapter 7, Section 7.3).

It is also important to clarify at this point what I mean by: 'predominantly positive' and 'predominantly critical' views. These 'labels' do not straightforwardly refer to a participant being an ultimate critic or proponent of the official approach to school evaluation. In other words, predominantly positive views do not imply that the participant was not critical of certain

aspects and, similarly, predominantly critical views do not imply that he or she took, at all times, a critical stance in his/her assessment of official approaches to school evaluation in England. Whether interview transcripts point more to positive or critical views may have also arisen from my interview strategy, my prompts, and probing questions. For instance, once I heard positive views in relation to the four *pillars*, I would ask the participants whether they saw any problems related to the *pillar* (see Appendix 2). I followed the same strategy with the participants who expressed critical views. For example: 'I can see you are critical of league tables. I wonder, do you consider they brought any benefit at all?' The result of such probes was that, whilst the participants with 'predominantly positive' views argued how the four *pillars* drove improvement in education, they also occasionally noted that previous approaches were inappropriate or led to 'unintended' consequences (e.g., Eduardo). The current approach, however, rarely received critique by participants with predominantly positive views, because they argued that either the approach evolved for the better or issues were arising in individual cases. In a similar vein, although participants with predominantly critical views argued how the four *pillars* were either a 'misconception' (Cooper) or caused negative consequences, they occasionally also referred to some benefits of these approaches. For instance, holding schools to account was seen by these experts as vital, with benefits such as ensuring that public money was spent well. However, the current approach, which considers accountability as a key driver of improvement, received a strong critique from these experts. The table below summarises the participants' tendency to refer to the official approach to school evaluation in England since 1988 either in a predominantly positive or predominantly critical light.

|  | Predominantly positive views | Predominantly critical views |
| --- | --- | --- |
| Policy actor | Bella, HMI inspector, policy adviser<br>Neil, Policy maker, policy adviser | Cooper, Local authority manager<br>Dennis, HMI inspector, researcher |
| Policy influencer | Eduardo, Researcher in school improvement, Ofsted Researcher | Nora, Researcher in assessment<br>Gordon, Researcher in inspection<br>Orion, Researcher in evaluation |
| School practitioner | Kent, Principal, Ofsted inspector | Kelvin, Headteacher<br>Torr, Teacher<br>Ned, Headteacher<br>Felicia, Deputy head, teacher |

**Table 4.4:** Overall views on the official approach to school evaluation in England

Some patterns emerging from the data summarised in Table 4.4 above indicate that *school practitioners* who participated in this study expressed predominantly critical views. Although

the participant sample was relatively small, the only *practitioner* who expressed predominantly positive views was Kent, who also trained as an Ofsted inspector. Similarly, the only *policy influencer* who expressed a predominantly positive view was Eduardo, who worked with *policy actors* as a researcher at Ofsted. Their involvement in the work of Ofsted may have given them insights that other *policy influencers* and *school practitioners* may not have been party to. This distinction guided my further data analysis. I then explored the benefits of the official approach and the problems arising from this approach cited by participants (see Chapter 6), as well as alternative approaches to school evaluation (see Chapter 7).

After I organised and synthesised them, I tried to understand what all that means, and I started writing up the findings. In other words, I started to tell the story of the data which sought to 'provide a concise, coherent, logical, nonrepetitive, and interesting account… within and across themes' (Braun and Clarke, 2006: 23).

I was also asked by some participants to double-check with them my analysis before publishing the research. I would argue that this is a crucial step enabling the researcher to build rapport with participants. Further, as suggested for sensitive research by Cohen et al. (2007: 125), these kinds of precautions would eliminate the feeling of being 'misrepresented by the research'.

Producing the report was a challenging process and my experience is in line with Harvey's (2010: 13) suggestion that interviewing elite members in a firm, unlike other employees, may mean that 'they are arguably more likely to represent the position of the firm rather than their own individual viewpoint.' Although I feel that my expert participants appeared to express their own views rather than those of their institution, some of them answered my questions by referring to theory, literature, or published reports. This could be because they did not want to disclose their personal, or institutional, view. As suggested by Harvey (2010: 18), for such answers, I did 'cross-checking and triangulating different kinds of evidence within the same interview' to verify the data.

## 4.6 Ethical considerations during and after data collection

In the stages of conduct, reporting and dissemination of findings, I took various precautions not only for a 'neat, clean, tidy, unproblematic and neutral process' but also 'to regard it as shot through with actual and potential sensitivities' (Cohen et al., 2007: 131). As suggested by Cohen et al. (2007: 127), 'the field of ethics in sensitive research is different from ethics in everyday research'. They suggested that the precautions cited in the literature on research

methodologies are however not 'universal' and local research needs may force the researcher to find out some immediate and practical solutions. Therefore, I recognise that the strategies that work for some researchers may not be effective for others.

These precautions were not only to do with protecting my participants and myself but also the participants' current or previously affiliated organisations. These considerations are expected to make a positive contribution to the 'consequential validity of the research' (Cohen et al., 2007: 132). For instance, I refrained from discussing some issues that either my participants or their affiliated associations might face. As suggested by Cohen et al. (2007: 124), 'rather than barring the research altogether, compromises may have to be reached in sampling and access'. In potentially problematic situations I prioritised the confidentiality of the participants. For instance, one participant shared a personal conversation s/he had with a high-level official at Ofsted, but I removed this conversation from my data as this could cause some speculation. Sharing this data was less important than protecting the confidentiality of the participant and his/her affiliated organisation. Also, one participant asked me to redact some of the words s/he said because this was: 'a highly controversial statement that might bring my organisation into disrepute'. I therefore immediately removed this statement as requested. In the process of checking interview transcripts, other participants also indicated which information should be removed and, again, I acted on their requests. As Cohen et al. (2007: 120) point out, 'what appears innocent to the researcher may be highly sensitive to the researched or to other parties' and, furthermore, ethical issues are felt 'mostly sharply if the research risks revealing negative findings' and 'researchers may not wish to take the risk of offending the powerful' (Cohen et al., 2007: 126).

## 4.7 Validity, reliability, and generalisability

This section explains how the issues of research validity, reliability and generalisability were approached in my study. My position on school evaluation as socially constructed inevitably affects the claims of validity, reliability, and generalisability of this research because social constructionist researchers are not interested in 'playing the truth game' and, instead, support ideas as 'possible resources' that social actors use when engaging in social practices (Gergen 2009: 160). Even though some researchers (Stake, 2005: 454) argue that the researcher's honesty and rigour can validate claims to new knowledge, the researcher's awareness of subjectivity and deliberate effort to question and disconfirm her own interpretations and gather depth data entail that I approach the issues of validity and reliability of this research with

caution.

Rather than justifying whether this research is valid, reliable and generalisable, I prefer to focus on explaining how this research could offer some generalisable insights. Because the knowledge about schools gleaned from different approaches to school evaluation is also socially constructed, the participant sample matters and I have, therefore, sought a 'balanced' participant sample. I have also sought participants' feedback and approval of interview transcripts. When analysing data, I have attempted to be reflexive and aware of how I positioned myself as a researcher and how subjectivity and bias might interfere with the processes of data analysis and reporting. To be transparent about the research process, I have also shared the challenges I faced when planning, collecting, analysing, and interpreting data in Sections 4.4 and 4.5 above. Expert interviews conducted and analysed in a rigorous, transparent way, may shed light on the benefits and problems arising from particular approaches to school evaluation. However, the findings from expert interviews should be read like all social science research findings. The findings from this study cannot be claimed to be generalisable to a broader population of educational experts or education professionals but instead, they can be seen as presenting a more, nuanced, complex picture of the diverse views and positions on the official approach to school evaluation in England and recommended alternatives. The ideas offered by experts who participated in this study could work as a possible resource informing improvement to school evaluation in English education or other education systems, if read carefully and approached with caution.

## 4.8 Conclusion

This study has focused on approaches to school evaluation in England since 1988, with a view to drawing implications for how school evaluation could be improved, within the English context, as well as education systems which rest on similar *pillars* as those that are now well established in the English system. To collect empirical data, I utilised the research method of expert (elite) interviews and conducted interviews with purposively selected educational experts who have had extensive experience and knowledge of approaches to school evaluation, both in the UK and internationally. My participant sample included highly esteemed academic researchers in the field of evaluation, policymakers and policy advisers, experts in school inspection, as well as experienced school practitioners working in leadership positions.

In this chapter, I elaborated on the processes involved in planning and carrying out my research, from the initial stage of research design to expert interview strategies and the analysis of data.

I have also discussed the preparation for expert interviews as an international female junior researcher and explained how I analysed the data, starting with coding in NVivo-12 and progressing to a thematic analysis that offers insights into expert participants' views on approaches to school evaluation. I embedded ethical considerations into planning, conducting, and reporting the interviews, as ethical issues mattered at every stage of my research, because of the sensitive nature of the expert interview. I would argue that conducting an expert (elite) interview necessitates looking at the conditions under which 'sensitivity' might arise within each step of the research process. I then considered issues related to the validity, reliability, and generalisability of this study, linked to my position as a social constructionist researcher. The following chapter introduces the experts who participated in this study, in order to enable the reader to 'meet' these experts and get a deeper understanding of the sources of their knowledge and education expertise.

# CHAPTER 5: EXPERT PARTICIPANTS AND THEIR PROFESSIONAL ROLES

## 5.1 Introduction

As explained in Chapter 4, the participants were purposively selected as educational experts with extensive professional experience and knowledge gained through years of work at different levels of the education system. 'Expertise' has been defined in this study in terms of 'knowledge' and 'experience' of the participants rather than 'status' and 'position' that are deemed relevant in some research using elite or expert interviews (Van Audenhove, 2017). Therefore, the participants who had in-depth expertise and knowledge (as discussed in Chapter 4) were deemed to be experts. Based on the analysis of answers to semi-structured interview question 1 (*Would you please tell me about yourself and your roles as an educational expert?* See Appendix 2), the participants' experience and professional roles in education were not clear cut. This is because the majority of the participants in this study (11 out of 15) started their career in education as teachers before moving on to other roles. This chapter introduces educational experts who participated in this study and discusses three categories of educational expertise (Section 5.2) developed by the participants, based on their roles in the education system: *policy actors* (Section 5.3); *policy influencers* (Section 5.4) and *school practitioners* as *policy implementers* or 'policy takers' (Gunter and Forrester, 2009) (Section 5.5).

## 5.2 Categories of educational expertise

If we define expertise in terms of 'knowledge' and 'experience' rather than 'status' and 'position' (Van Audenhove, 2017), then the knowledge and experience of the participants in this study rested on three main types of roles in which they worked throughout their professional careers. These are summarised in Table 5.1 below. 'P' refers to previous roles and 'C' to the current role, i.e., their role at the time when the interview data were collected.

| Category | Policy Actor | | | | | | Policy Influencer | | School Practitioner | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | HMI and Ofsted inspector | Policy maker | LA management | Policy Adviser | QA of inspection | Adjudicator | Researcher | Author | National Leader of Education | National Leader of Governance | School senior Leader | School teacher | School governor |
| Richard | P | C | P | C | | | | | | | P | P | |
| Cooper | | | P | | | | C | | | | P | P | C |
| Dennis | P | | | P | | | P | C | | C | P | P | C |
| Gabriella | C | | | C | C | | | | | | P | P | |
| Bella | P | P | | P | P | C | | | | | P | P | |
| Neil | | C | | C | | | | | | | | | |
| Nora | | | | | | | C | C | | | | | |
| Gordon | | | | | | | P | C | | | | P | |
| Eduardo | | | | P | | | C | C | | | | | |
| Orion | | | | | | | C | C | | | | | C |
| Kelvin | | | | | | | | | | | P | P | |
| Torr | | | | | | | | | | | | C | C |
| Felicia | | | | | | | | | | | C | C | |
| Ned | | | | | | | | | C | | C | P | |
| Kent | P | | | | | | | | | | C | P | |

**Table 5.1:** Current (C) and previous (P) roles of participants

As can be seen in Table 5.1, the expertise of the participants was developed through diverse professional roles and activities which they undertook. Therefore, assigning a single type of role and related expertise to each participant was challenging. For example, Richard worked as inspector, policy maker, senior school leader, and schoolteacher, as well as in management roles in local authorities and inspectorates. He has worked in the evaluation field as a policy maker for the last fifteen years in various countries and emphasised that he could talk either from a 'recipient' or Ofsted inspector perspective:

> I could offer my views on this from two point of views. So, firstly, as a 'recipient', as a school Principal in the school, I found the inspections actually to be fair, rigorous, well-managed and accurate. The teams that were assigned to my school were well qualified and sufficiently experienced… Regarding the actual inspection itself, once I was deployed as an Ofsted team inspector, I had more concerns with, and I saw a significant variability in the quality of inspections and the depth of quality assurance of the process. (Richard)

Three broad categories emerged encapsulating the participants' key domains of expertise,

although a clear-cut separation was not possible. Consequently, the role performed by the participant at the time of data collection was deemed to have a greater bearing on the emerging categories than the length of experience in other roles. These categories include*: policy actors, policy influencers* and *school practitioners.* How these categories were identified is explained and illustrated below.

## 5.3 Policy actors

*Policy actors'* domain of expertise was mainly based on advising, designing and/or monitoring the implementation of policies. The specific professional roles within this category included local authority officers, Ofsted Inspectors, HMI Inspectors, senior officers working for the DfE or Ofsted. This section presents a 'potted history' of participants whose expertise was aligned to that of *policy actors*.

Richard started his professional career as a teacher in the early 1980s and became a principal of two schools in the early years of the establishment of Ofsted in England. His school roles lasted around two decades, whilst the remaining two decades of his career were spent in the *policy actor* roles. The *policy actor* roles varied from the design of new inspection regimes and policies of various countries to inspecting schools for both Ofsted and the Scottish HMI. While reforming inspection regimes abroad, Richard mostly 'borrowed' from the Scottish Inspectorate policies and methodologies (removing market-driven models from these countries). He also carried out management roles as a *policy actor* at both local and government levels as Director of Education and Director of Inspectorate in various countries. At the time of data collection, he still worked as a policy maker and as a Chief Inspector for a government of another country (for ethical reasons the details of this role are omitted from this thesis).

Cooper started his professional career in the early 1970s, as a teacher. At the time of data collection, he worked as a school governor. However, most of his career has been devoted to designing policies from the early 1980s. When the Education Reform Act (ERA) was introduced in 1988, he was a managing officer in a local authority tasked with introducing the reforms that come along with ERA and, from 1992, supporting schools with Ofsted inspections. Throughout his career, he worked for four different local authorities in management posts that included Chief Education Officer and Assistant Chief Education Officer. As he pointed out, he 'lived through the history of the whole thing - the old system, the change from the old system, and the current new system of course' (Cooper).

Dennis started his professional career as a teacher in the 1960s and then undertook head of department and deputy head roles. His professional career continued in the roles of a researcher and lecturer besides being a frequent contributor to the national press for over fifty years. Apart from the *school practitioner* and *policy influencer* roles, he also undertook a range of roles in the *policy actor* category starting from the 1980s until the mid-1990s. Before and after Ofsted he worked as an HMI Inspector and undertook some advisory roles for the DfE. When working for Ofsted, he undertook various roles from staff inspector to Ofsted's specialist adviser. At the time of data collection, he worked as a National Leader of Governance and focused on accelerating school improvement through his support.

Gabriella undertook a headship role for over a decade and also worked for Ofsted. At the time of data collection, she worked as a school inspector in a devolved education system, carrying out various roles, from the quality assurance of school inspections to adviser roles. For reasons of confidentiality, I am unable to specify in which of the three devolved systems she has worked.

Bella started her professional career as a teacher in the mid-1970s. When the ERA 1988 was introduced, she was a new headteacher in her first year of headship. After working as a headteacher in two different secondary schools in deprived areas for fifteen years, she continued her career as an HMI inspector for almost two decades in England, from the 2000s, focusing in particular on working with schools under special measures. In this role, she represented Ofsted at international conferences. As an Ofsted inspector and adviser, she contributed to the quality assurance of more than ten thousand Ofsted inspections, carried out every year across the system. She also worked on the development of one of the Ofsted inspection frameworks in the 2000s. At the time of data collection, she worked as a school adjudicator for the DfE.

Neil undertook the *policy maker* roles in central government and its agencies and was responsible for a range of strategic and operational tasks. These included developing Ofsted inspection frameworks and consulting schools on these frameworks. For ethical reasons, it is impossible to provide further details of Neil's work.

## 5.4 Policy influencers

*Policy influencers'* domain of expertise was derived mainly from seeking to influence policies through research, policy analysis and/or national campaigns. The specific *policy influencer*

roles include think tank researchers, academic researchers, union members, as well as members of professional associations.

Nora started her professional career in the late 1980s working with students at schools for charities and non-governmental organisations on the development of education programmes as a facilitator. She continued her career in education as a researcher and her expertise is focused on teacher professionalism and student assessment. In her early career, she worked in a research team as a senior research officer for an 'awarding body' ('exam board'). In this role, she visited schools to interview children and spend time observing students and teachers in the classroom and conducted national studies about examinations and tests and their impact on students. Nora emphasised that she observed a significant change from the 1990s with teachers 'becoming quite anxious' about accountability and increasingly preoccupied with 'criteria', 'statistics', and 'hard' assessment data. This made her ask the following questions: 'Do we all mean the same thing when we all talk about… 'grade A'? Can we be absolutely sure that we know what it is?' Although she described the time before the 1980s as 'dark times in England' because of the differences between individual schools and teachers, the 'split' between 'what was going on in school' and 'how a good school, a good teacher or a good student is measured' enticed her to work as an academic and researcher in the field of assessment. She also took a part in international education reform projects by working for the OECD.

Gordon started as a teacher in the 1960s. Over five decades, he worked as a researcher in England and Scotland. One of his long-term jobs was training new teachers. His work as a *policy influencer* led to a range of policies being introduced, modified or withdrawn in response to his research and campaigning. His expertise stems from his knowledge of the history of school evaluation in England, with his research impacting on policy and Ofsted inspection. He also researched school inspection systems in other countries. In order to ensure confidentiality, I am unable to give more detailed examples of Gordon's work.

Eduardo is a researcher who has worked closely with policy makers. His expertise stems from his knowledge and research on school evaluation systems in a number of countries, as well as his research on school accountability, improvement, effectiveness, and inspection. At the time of data collection, he was working as a researcher at Ofsted, with a remit for informing Ofsted frameworks and policies. To ensure confidentiality, details of Eduardo's research and professional roles are omitted.

Orion has worked as a researcher in educational evaluation since the mid-1970s, collaborating with leading evaluation theorists in the UK and internationally. His expertise thus stems from his in-depth knowledge of evaluation theories and practices. He has also worked as a school governor. He is one of a few participants in this study (similar to Eduardo and Neil) who did not start his career in education as a *school practitioner*.

## 5.5 School practitioners

*School practitioners'* domain of expertise is derived from enacting policies at the school level. The specific roles within this category include membership of the senior leadership team (i.e., headteacher, principal, deputy head), experienced schoolteachers or governors.

Kelvin worked in schools throughout his professional career. As a headteacher of two schools for 12 years, he received all four Ofsted grades. His first school was a local authority school and the second one was an academy. He considered his engagement with evaluation bodies to be relatively high: 'I've had a lot of accountability. I don't think you'll find many people who had more.' He received eleven inspection visits in total: eight by Ofsted, three by the DfE (Department for Education), MAT (Multi Academy Trust) and RSC (Regional Schools Commissioners). Nine out of these eleven inspections were in his last five years (six of them were from Ofsted). Being 'over-evaluated and underappreciated' were the reasons for his early retirement. His final school was a newly set up academy, 'a challenging school in a deprived area, with haemorrhaging students… there was a lot of government expectation that being an academy would lead to rapid improvement'. In a few weeks after he started work, the academy received its first Ofsted inspection and was graded as 'inadequate'. In the last inspection before he left the school, it was graded as 'good' and became oversubscribed for three years in a row. However, 'other people have done less than that and got knighthoods', and his contribution 'is not picked up… because of the timing of the Ofsted schedule'. The amount of inspection pressure he experienced made him reflect that headship: 'is not an attractive profession to people. I left early. I wanted to go for another two years. I could not. I didn't want another two years of being over-evaluated and underappreciated.'

Torr worked as a subject teacher for quarter of a century and started just after the establishment of Ofsted. He worked for two rural schools. As he observed, he has not 'seen a lot of different schools' but has seen 'lots of inspections in the same context [in his second school]'. He was in charge of his subject for around 15 years and at the time of data collection also served as a teacher governor.

Felicia started her professional career as a teacher two decades ago and worked in both academy and local authority schools. She also worked as a teacher abroad, which gave her a different perspective on education in a country which does not have league tables, and teacher targets. At the time of data collection, she held two roles at her academy as a teacher and assistant principal. In her senior school leadership role, she is responsible for the progress and achievement of students. In this role, she makes sure the students are 'hitting… progress measures' and 'the particular vulnerable groups are making the progress that they should'.

Ned was a newly qualified teacher in the year the ERA 1988 was introduced. He undertook his first senior leadership role as a deputy head of a school in the mid-1990s. With almost two decades of headship experience at two different schools, he also works as a lead headteacher in his local authority. His second and current school, graded as 'good' by Ofsted, a large local authority school with over a thousand students, is a National Support School (NSS) and he is a National Leader for Education (NLE) to deliver school improvement support to other schools.

Kent's professional career started as a teacher at the end of the 1990s. In five different schools, he undertook various senior leadership roles such as head of department, head of year, assistant principal, deputy headteacher, vice principal. At the time of data collection, he worked as a principal (with 'three' years' experience) of an academy in a Multi Academy Trust with eight schools in it. He is a qualified Ofsted inspector but has not practised yet 'because it just doesn't work along with the school.'

## 5.6 Conclusion

Overall, the expert informants' professional expertise was derived from the experience of working in education which ranged from almost two to over five decades. Nine participants were in post in 1988, the year in which Education Reform Act (ERA) was passed. Four participants started to work at the time when Ofsted was established in the mid-1990s. Being the witness of, and participant in, historical milestones in the development of school evaluation such as the 1988 ERA and the establishment of Ofsted gave the participants privileged knowledge and understanding. However, in addition to their historical knowledge, their expertise also extended to theoretical and practical knowledge. Four participants have achieved the role of a professor. Amongst school practitioners, Torr and Felicia had postgraduate and Master's degrees, in addition to National College qualifications. Eleven experts undertook the roles of senior leadership, including headship or served as governors of schools or groups of schools. Five experts took policy advisory role and three worked for DfE. All informants had

experience in the English education system and seven also have had knowledge of other systems (i.e., New Zealand, Dubai, United Arab Emirates, Berlin jurisdiction in Germany). The professional experience of experts who participated in this research was diverse and included roles in policy development as *policy actors*, researching and policy critique as *policy influencers*, as well as being at the receiving end of policies and approaches to school evaluation as *school practitioners*. Therefore, 'expertise' in this study refers to professionals who have historical, technical, theoretical, and practical knowledge of school evaluation. How this knowledge was drawn upon to assess the evolving approach to school evaluation in England since 1988 provides the focus of the following three chapters.

# CHAPTER 6: EDUCATIONAL EXPERTS' ASSESSMENT OF THE OFFICIAL APPROACH TO SCHOOL EVALUATION

## 6.1 Introduction

This is the first of two chapters of this thesis focused on the discussion of empirical data collected through semi-structured interviews with educational experts. The chapter addresses research question 1 pertaining to the participants' views on the official approach to school evaluation. The participants and their professional experience and roles as educational experts have been presented in Chapter 5. As a social constructionist researcher, I was interested in how the participants talked about their professional experience and expertise and whether the range of their expertise and experience had a bearing on their views about approaches to school evaluation. Since the participant sample was relatively small, I was unable to find clear patterns that would link the participants' 'potted history' to their views on evaluation. The discussion of the participants' views on the changes to the official approach to school evaluation in England post 1988, with the benefits and problems arising from this approach, are discussed in Sections 6.2 and 6.3 respectively. As explained in Chapter 4, Section 4.5, the analysis of interview data pointed to a strong polarisation of experts' views and the same *pillars* of school evaluation were assessed as either the cause of problems or the source of strengths in the education system, depending on whether the experts were predominantly supportive or critical of the official approach. The data also showed that being a critic does not mean being 'radical' in offering alternative solutions. I elaborate on this finding in Chapter 7, but it seems that positively orientated participants, as well as some of the critics of the system, take the current system for granted. Taking a system for granted may not necessarily be a problem when the system works well. But when it does not, a positive orientation may hinder the drive for change necessary to improve the system. In the following two sections, I present the findings about the benefits of the official approach to school evaluation (Section 6.2) and problems arising from this approach (Section 6.3).

## 6.2 Benefits of the official approach to school evaluation

This study found out that the benefits of the official approach to school evaluation since 1988 cited by participants included: national standards and frameworks, accountability to Ofsted, high stakes as a tool for 'bringing people into line' (Kent) and transparency to parents by providing the 'right information' (Bella, Eduardo) through Ofsted reports and school league

tables. These were cited as beneficial by experts who held predominantly positive views on the official approach because they were seen as driving improvement at the school and national levels.

### 6.2.1 National standards and 'quality' frameworks to 'make people act'

An important benefit cited by participants was captured by Eduardo as the explanation that national standards and frameworks 'make people act' in a 'right direction'. Whilst before 1988, there was ambiguity around the meaning of 'quality' because of the lack of national standards, post 1992, Ofsted undertook an important role by introducing their 'clear' frameworks (Bella). Bella pointed to the differentiation between local authorities before 1992 and emphasised how 'good' local authorities would know 'what was good and bad' in their schools. However, the lack of national standards meant that 'there wasn't necessarily a national standard of what an outstanding school or an outstanding lesson looked like'. Bella described Ofsted as 'all about national standardisation' and argued that 'the core of what we do', i.e., teaching and learning, improved 'tremendously' since she started teaching in the 1970s, due to Ofsted's standards. In a similar vein, Kent argued that, with the absence of national standards, success relied on individuals, and this led to differentiation in school quality pre-1992. Post 1992, Ofsted became a 'fundamental reason to the shift in improvement':

> Pre 1988, it was like a wasteland… You've got a good school; you've got a good school. If you didn't, you didn't. If you were a headteacher, you were autonomous. You might have local authority visits but nothing, nothing special. So, the introduction of the inspectorate was fundamental. I think, it's an absolutely fundamental reason to the shift in improvement. We wouldn't be where we are now if we hadn't had it. (Kent)

The participants also explained how the standards enable improvement through 'clear' frameworks (Bella). For instance, Eduardo, as a representative from Ofsted, referred to the 'mechanisms' through which standards work. He explained that 'certain frameworks and parameters' create 'expectation' in the system and people 'act' on them and this drives 'indirect' improvement or 'enforced improvement' (he used both phrases) at the school and system level. Referring to the recent Ofsted focus on 'curriculum and quality of education', Eduardo explained how their framework is a 'force for improvement within the system'. As he put it:

> we have to look at the system as a whole and say 'OK, these things are going well. These things are not going well'… that does allow us when necessary to re-steer the system in a new direction. So, for example, we found that schools are concentrating too much on examinations, we come through our inspection system trying to steer that in a different direction… by producing particular frameworks. For example, we put a debate in motion, and we get schools to think about stuff. An example is the new inspection

framework where we have said: 'OK, we are going to focus much more on the curriculum and quality of education'. Now of course schools have always thought about curriculum to some extent. I think, certainly what you can see now is that there is a lot of debate suddenly within the system within schools around curriculum. So, also by setting certain frameworks, certain parameters that creates expectations. So, you create a quality framework and that, that in itself makes people act, make schools act, make governors act, make local authorities act etc. So, also in that way, you can be a force for improvement within the system. (Eduardo)

Eduardo's explanation resonated with Kent's argument who, in his role of *school practitioner*, said that 'if you are driven by standards, your standards do improve'. He argued that Ofsted's focus on outcomes data led to school improvement:

Ofsted… has been one of the motivators to ensure schools do improve… And I think, there is substantial evidence to suggest that the focus from Ofsted on data has improved schools. You know, there isn't any doubt that has had an impact. (Kent)

Overall, the national inspectorate with 'clear' standards set out within a framework was seen as the most important element of post 1988 reforms because standards made it clear what was expected and what 'outstanding looks like' (Bella). Standards, frameworks, and evaluation schedules were presented as useful for 'setting expectations' and providing a 'steer' for people to 'act' on (Eduardo). In this mechanism, Ofsted decreased the differences between schools, compared to the times before 1988, improving the quality of education overall. These experts emphasised the importance of 'reliable' (Eduardo, Neil), 'valid' (Neil), 'fair' (Eduardo), 'objective' (Bella) and 'consistent' (Neil) judgements, based on standardisation. They also highlighted the importance of professionalism and expertise of inspectors in reaching their judgement. The 'autonomy [of inspectors] within a framework' (Eduardo) was deemed to be an important balance to be achieved by Ofsted inspectors so that standardisation could bring the benefit of improvement to individual schools and the whole system. The autonomy of inspectors is discussed further in Chapter 8.

### 6.2.2 Accountability to Ofsted

This study found out that an accountability system which rests mainly on Ofsted is beneficial to drive school improvement. Accountability was seen as evolving as a result of the Education Reform Act (ERA) 1988. For instance, Eduardo and Bella argued that accountability existed before 1988 but was 'probably weaker' (Eduardo) through schools being held to account by local authority advisers and inspectors. However, Neil and Kent argued that the system in these days was 'without accountability' due to insufficient local authority inspection. Success depended on 'luck' (Kent). Importantly, these experts talked about the accountability of

schools to Ofsted rather than to the government or other stakeholders. The function of accountability was seen as putting external pressure through inspectors' judgement and steering the school practice in the 'right direction' (Neil), as elaborated on further in Chapter 8. This section focuses on how these participants problematised the time before 1988 and how they argued that accountability has been a key driver of school improvement post 1988, with Ofsted as the key agency ensuring that schools are held to account.

Specifically, the participants argued that quality improved through holding schools to account and that, in addition, accountability also reduced the difference in quality between schools in England. Neil, for instance, argued that if 'there is no external spark, it gets worse'. He suggested that one looks at the English education system prior to ERA, particularly in the 1970s, 'to see what a system without accountability is.' For him, 'the privileged' children in the private sector used to get a better deal than children in the state sector and, after 1988, accountability worked particularly for those previously 'left to flounder'. Although 'lots of school improvement can happen without accountability', Neil also argued that being accountable to 'somebody external' is important because 'somebody needs to see', especially in the case of the inadequate schools in the system. Neil explained that 'accountability nudges behaviour into a certain pattern' and that 'knowing that you'll be asked about it keeps you on your toes and keeps people who are good doing the right things'. Interestingly, Neil's idea of the benefits of accountability was framed within negative terms, in the sense of schools being held to account by someone 'external to us whom we hate':

> Lots of school improvement can happen without accountability, lots of it. But we all need somebody external to us whom we hate to make us get out of bed in the morning or to make us learn something that we don't want to learn, or to make us make a change that we don't want to make... So, 86 percent of schools in the country are good or better according to us. Four percent of them are inadequate. There are four percent of schools out there where I would not want my child within a country mile of that school. Somebody needs to say it. Somebody needs to see it, and somebody needs to say it. And that's what accountability does… Now, we're not liked, we're not popular for doing that. But we're not there to be liked, popular. We're there to create the conditions where the education would turn right for the kids who are in that school. Because, too often, that kid will go through the whole school in a terrible environment, learn nothing and be sent back for the rest of their lives. And, so, to me, that's the sharp end of what our accountability does. (Neil)

As can be seen from the point above about 'turning' education 'right' for the kids, Neil's rationale for introducing accountability was based on a 'deficiency model' of education, with children 'too often' learning in a 'terrible environment'. This suggests that the 'nudging' and 'keeping you on your toes' is necessary and, by extrapolation, without the 'sharp end' of

accountability, children would continue in a 'terrible environment':

> And I think, at the better end, accountability nudges behaviour into a certain pattern. Umm, sometimes intentionally sometimes unintentionally… knowing that you'll be asked about it keeps you on your toes and keeps people good, doing the right things. (Neil)

In a similar vein, Kent, who claimed that there was not accountability before 1988, argued that holding schools to account decreased the number of 'really bad' schools. For him, 'the real heavy accountability' came with Ofsted. As he put it:

> Before 1988, you know, before evaluation, there were some really bad schools, really bad. And I'm not saying there aren't still some really bad schools. But there are far, far fewer now than where there were. And I don't think, there's anybody that could argue that actually that hasn't been in large part due to the inspection regime that we have…You know, before 1988, there was nothing. I mean there was nothing. There was obviously nothing. There was nothing to make sure schools were doing the job… You know, nobody was accountable... Accountability means kids are doing better. If you're doing well, the kids are doing better. And I take that personally quite seriously... So, I don't have a problem with it. It shouldn't be the whole thing, but I don't have a problem with it. (Kent)

Kent's repeated references to 'nothing' before accountability was introduced, suggest that one important benefit of accountability is that it is a vital driver of school improvement. However, Bella and Eduardo stated that accountability existed before 1988, except that it was delegated to local authorities. For Eduardo, claiming that the pre 1988 days were non-accountable days is 'overstating the case' because 'schools at that time were managed by their local authority and local authorities did have local accountability mechanisms in place'. Importantly, these three participants seemed to suggest that 'real, heavy accountability' (Kent) came with Ofsted and accountability delegated to local authorities was not of the same 'weight' as accounting to Ofsted inspectors. Thus, Eduardo distinguished between 'weak' and 'strong' accountability and argued that some local authorities implemented 'stronger accountability and also stronger support for schools' because 'different local authorities did have local accountability mechanisms in place'. He noted that accountability is an important element to school improvement, although it is 'not only one' and argued that England introduced 'more rigorous' accountability in order to ensure 'sufficient' government oversight:

> At the same time, a movement towards giving schools more freedom and giving parents more choice, but also some of the things moved to make sure that there is sufficient oversight from government to enhance accountability measures. We of course see the publication of school performance tables happening and then to see Ofsted taking over from the old HMI system and introducing more rigorous accountability. (Eduardo)

Although Neil, Eduardo and Kent acknowledged that accountability improves schools, Bella

hardly used the notion 'accountability' and appeared to 'hedge' my direct question about accountability. Instead, she described how accountability is about headship responsibility because headteachers are 'accountable for everything that went on' in their schools. However, her view of accountability as a driver of school improvement appeared to be positive in her answer to another question. When asked to imagine an education system without accountability, she said that, in a system without accountability, practice 'becomes wishy-washy, poor practice, mediocre'. In addition, Bella strongly disagreed with those who argue that schools were not accountable prior to 1988 and explained that they are likely to misinterpret this because the reforms after 1988 were about national standards and 'not about accountability'. She described how, as a teacher, she was held accountable internally to the heads of department and headteachers and her schools were held accountable externally more to local authorities and less to HMI. The role of Her Majesty's Inspectorate was mainly to look at the quality of the system and these inspectors rarely visited schools:

> I started teaching in 1975 and I was, I was definitely accountable for what went on in my classroom. My head of department checked that I was doing what I should be doing. The headteacher of the school checked that the department was doing what they should be doing… also, although there was less inspection [HMI] obviously, there was more from the local authority. Because local authorities had teams of inspectors, and they called themselves inspectors rather than advisers - who would come in and support schools… You couldn't just go in and do what you like. (Bella)

The participants considered accountability to be an effective driver of school improvement because of its function to 'nudge the education system in the right direction' (Neil) through external 'neutral' (Eduardo) views. Eduardo gave a broader picture of how accountability works 'directly' and 'indirectly' to improve schools ('indirect' improvement occurs through the set standards, as elaborated in Section 6.2.1). He claimed that accountability 'directly' improves schools because people 'take action' on 'reports' and 'exam results.' In this regard, he argued that an 'external neutral party' can be sometimes helpful to show 'blind' weaknesses that schools could not see:

> And very often it is helpful to have an external neutral party, taking a look at that from national perspective. Because it is… you don't always… you don't know what you don't know. So, umm, in any organisation, you will be sometimes blind to some of your own weaknesses. (Eduardo)

These experts also explained how accountability evolved over time. For instance, Bella highlighted that accountability by school leadership changed 'for the better' and pointed out that this occurred particularly with the instigation of large Academy Trusts as 'there is shared accountability there'. Kent, Neil, and Eduardo focused on the changes to the accountability

measures. Kent for instance found Ofsted's role of evaluating the 'quality of education' as 'better' with the recent 2019 Ofsted Framework. Although the negative impact of outcomes-based accountability measures was associated by Kent with a 'perception that schools have become exam factories', he highlighted how the focus of Ofsted on 'outcomes' worked as a 'motivator' for the schools to get 'better results.'

> if you are held accountable on the basis of outcomes and other stuff… if you are driven by standards, your standards do improve. And I think, there is substantial evidence to suggest that the focus from Ofsted on data has improved schools… The negative impact has been… is the perception that schools have become exam factories. Actually, it is all about exams. And what one has seen as a result of that accountability from Ofsted is the removal of things like performing arts, visual arts, sports, RE and so on. And those have definitely suffered… outcomes for children have got better and that's not just happened through osmosis. It's happened at least in part as a result of the inspection and evaluation framework. I've no doubt about that at all. (Kent)

Neil and Eduardo highlighted the reasons for changes to accountability as a 'swing of the pendulum back' from 'excessive focus on data' (Neil). For them, this 'pendulum swing' was related to an increasing focus on school outcomes rather than quality and they claimed that this might be related to Ofsted's reliance on 'attainment data in terms of reaching its inspection judgments' (Eduardo). They pointed out that, in 1992, when Ofsted came into existence, the conception of accountability was established on two 'counterbalancing arms' (Neil). First, performance tables and pupil outcomes through the various national tests and, second, school processes through inspection. These two 'were meant to keep each other in balance' (Neil) as 'complementary' (Eduardo). However, Neil argued that this introduced 'ideas' into the system around performance targets, metrics, managing through data 'with the rise of managerialism in the 90s'. These were 'useful ideas' until they 'started to take over'. Thus, performance tables became 'disproportionately powerful', and Ofsted inspections started to reflect the school performance table. These 'two counterbalancing forces' turned into 'one mutually reinforcing force'. The focus was on 'trying to get the best results in tests but not thinking more widely about what children should learn'. Further, he argued that the 'national strategies' in literacy and numeracy unintentionally 'eroded teachers' role in thinking about the curriculum, thinking about pedagogy, increased the focus on data and removed the qualitative counterbalance'. Thus, from 2010-2011, 'the system started to chip away'. Therefore, particularly over the last two years, Ofsted started 'to swing the pendulum back' from an 'excessive focus on data'. He said that, with their latest 2019 framework, 'standards matter but you've achieved standards by thinking about the great curriculum'. Neil's 'evolution' of accountability was conveyed through the following 'broad-brush description', cited here in full because it provides a highly

evocative account of the 'pendulum swings':

> 1988, birth of the national curriculum and it follows on from the Callaghan's 'Secret Garden' speech of the 70s which I think is hugely important. I'd pick up the story in 92 because that's when Ofsted came into existence. And, at that point, the conception of accountability was to have two balancing arms. 'Performance tables' which are the accountability to parents about 'is the education in their school good enough and are their children getting good results?' And 'Ofsted' which was qualitative, which was 'what is it like in their school?'. And it seems extraordinary but an inspection in those days would use up to 44 inspector days and we'd crawl over every subject, every aspect of the school and it would be qualitative. And the two were meant to keep each other in balance.
>
> What then happened throughout the 90s was the… umm, two things. First of all, the Ofsted model in those days probably wasn't sustainable. Forty-four days crawling over a school for that long, our reports were not sharp enough, all sorts of things. So, so there had to be some kind of change.
>
> The rise of managerialism in the 90s… introduced a lot of ideas around performance targets, metrics, managing through data. And that was useful because education didn't have that concept before. But it started to take over. Umm, and performance tables became disproportionately powerful and there was a long process of moving away from Ofsted original role and making inspections smaller and smaller and smaller and smaller until ultimately in 2015. For a primary school, you'd have one inspector for one day. So, from 44 days down to one. And Ofsted inspections started to reflect a lot of the data. So, instead of having two counterbalancing forces, you had one mutually reinforcing force… that led to a system in which data was 'king', in which the focus was on tests, on trying to get the best results and tests but not thinking about more widely what children should learn.
>
> At the same time from 97, there was a focus on literacy and numeracy. And the numeracy area and literacy area in primary schools hopefully are… good initial conception and necessary. But the national strategies became an enormous monster that drove practice and said to teachers: 'Don't worry about thinking what you need to teach or how you need to teach, just use this material'. So, we eroded teacher's role in thinking about the curriculum, thinking about pedagogy, increased the focus on data and removed the qualitative counterbalance. And then, we spent a long time wondering why the gap wasn't closing more than most disadvantaged children were not catching up with their richer most prosperous peers.
>
> And then, since 2010-2011, we have the system starting to chip away that. And, strangely, maybe not strangely, over the last two years, Ofsted has been leading… the charge against that. So, our work for the last two years has been entirely about 'bringing the curriculum back into the centre of the educational conversation' and saying 'standards matter but you've achieved standards by thinking about the great curriculum… well taught that leads through to pupils knowing more as a member or being able to do more'. And that's where we are now. The education establishment is very split and polarised into progressivism and traditionalist forms and trends. And, so, the position that we're advocating is strongly resisted by some. And the real moment is going to be September when the new framework comes into place, when we see whether we start to swing the pendulum back. (Neil)

In a similar vein, Eduardo explained how accountability evolved from its 'initial intention' and

how student performance results and Ofsted inspections 'have grown closer together.' Ofsted then became reliant on attainment data for reaching inspection judgments, resulting in excessive emphasis on test results which led, in turn, to schools gaming the system. Therefore, Ofsted decided to 'reboot that a little bit and look at evolving our inspection framework to focus less on that test and attainment data and more on the processes that we would have to look at in the first'.

Overall, the argument that the local mechanisms to hold schools to account pre 1992 were 'weak' (Eduardo) or insufficient due to the lack of 'real, heavy' accountability (Kent) resonates with the arguments by policymakers and supporters of the official approach to school evaluation discussed in Chapter 2 (Gilbert, 2012). The role of HMI was focused more on the quality of the system than on individual schools. Before 1988, there were some really 'bad' schools (Kent, Bella) and the key benefit of new accountability mechanisms through performance tables and Ofsted inspection was that they worked as an effective driver of school improvement. The focus on outcomes drove improvement and 'there's no one that can argue that children getting better results isn't a good thing' (Kent). However, school performance results (reported in 'league tables') and Ofsted inspections 'have grown closer together' (Eduardo) although, initially, they were two different pillars in the official accountability mechanism. Ofsted then became reliant on attainment data for reaching inspection judgments, with the unintended consequences of the strong emphasis on test results including 'gaming the system' (Eduardo), 'eroding' of the teachers' role (Neil) and 'excessive focus on data' (Neil). This, in turn, necessitated a 'pendulum swing' led by Ofsted (Neil). The evolution of accountability reported by these participants was evaluated as change for the 'better' (Kent), from outcomes-based accountability to the quality of education-based approach, particularly with the introduction of the 2019 Ofsted Framework. An important issue on which there was a consensus amongst these participants was to do with trust and a belief that a system based on trust rather than accountability cannot work. This is discussed further in Chapter 8.

### 6.2.3 High stakes accountability as a tool for 'bringing people into line' and its unintended consequences

The experts who expressed predominantly positive views on the official approach to school evaluation argued that high stakes accountability was needed in England, although it may 'create distortions' (Eduardo) and decrease the quality of information gathered through accountability mechanisms. Eduardo argued that 'schools will naturally try and do whatever

they can to get the best possible outcomes in terms of accountability'. Because the system with high stakes creates pressure, particularly on the headteacher, this can 'lead to those issues of gaming, unintended consequences.' On the other hand, he emphasised that high stakes are important because heads in England have 'high autonomy':

> England has a high stakes accountability system as you know but also has a great deal of freedom for schools alongside that as well. So, headteachers for example can take decisions about things that for a lot of other countries would be unthinkable like budgets, hiring teachers, tweaking curriculum etc. So, I think, those things go together. Say in other systems where you do not have that freedom at individual school level; you typically have less accountability in terms of maybe assessments or you have where you have inspection system, they are weaker… I say, we are quite on one end of the continuum in terms of this combination of high levels of autonomy and high levels of accountability for schools. (Eduardo)

Neil, like Eduardo, acknowledged unintended consequences of high stakes but also highlighted their importance. He argued that 'when we [Ofsted] speak, people listen, but they don't always hear' and this is connected to high stakes. Thus, there will always be 'unintended consequences' naturally and 'as long as we exist and as long as we have the consequences attached to our judgments, we will drive those unintended consequences.' As he explained:

> because we have high stakes that creates distortions. And because when Ofsted speaks everybody 'jumps' and they may not jump in the right direction… until three years ago, we published good practice examples on our website. Government always wanted to present good practices. We published a case study about a school that did marking with seven colours of pen. Thousands of schools in the country bought seven colours a pen and made all their teachers do that obsessively. And we didn't ask them to… but we just said 'Oh, look, someone's doing that well and that works in that context.' So, when we speak; people listen, but they don't always hear. So, there are unintended consequences of whatever we do. And that connects to the stakes around us… And we've done a huge amount of work 'myth busting' on Twitter and all that kind of stuff. But we'll never get rid of that problem. And that's, that's one of the great regrettable things about who we are. (Neil)

Kent and Bella spoke about unintended consequences as an individual (rather than systemic) matter. Kent for instance described accountability as 'a double-edged sword' and acknowledged how 'some' schools are 'ditching subjects' which are not part of high stakes accountability. However, he emphasised that he does not have 'that panic' because he is 'confident of what we're doing'. In relation to the pressure of accountability, he argued that 'people are big enough to cope with that' and further explained that staff stress is related to the headteacher's attitude. As he put it:

> I've never been involved in an inspection where the team has been out to get you. I've never had that... You hear stories, I'm sure. Ofsted inspection teams or inspection is going to schools and being a horror story. You know, the terrible things they've said or done or whatever. I've never experienced that… All the inspection teams have been open and transparent and whatever… I think, what Ofsted are looking for is completely

fair. We've all read it back-to-back; and, yes, it's a reasonable document. You know, you can't make fifty thousand headteachers happy… But you can't please everyone. You won't please the people whose inspection experience has been poor. But those persons whose inspection experience has been positive, you're going to get a reasonable response… And the experience of being inspected is nerve wracking, it's frightening, it's all of those things. But you have to get over that a little bit… I think, we all big enough to cope with that. (Kent)

Similarly, Bella argued that headteachers' stress is about 'individual' situations and people who do not want their weaknesses to come into the open. Because Ofsted inspectors are 'objective' professionals, their reports will reflect what they see at the school. She also suggested there is a lot of 'hype' and 'myths' about Ofsted and it should not be 'perceived as beating the schools with a big stick' because it is 'a professional development model'. A 'sensible school embraces Ofsted' and 'uses it as a tool for their own improvement':

And, so, to say negatives, I don't know. I think, a lot of it is to do with the individual people who are involved both in the school who perhaps are a bit frightened of Ofsted and they don't need to be or of inspectors. You know, like in every profession, I'm sure, there are lots and lots of very, very professional inspectors, but there are going to be one or two who perhaps aren't as caring or as professionally based as they might be… And, if you read the evaluation schedule, it is a professional development model… It isn't a beating them with a big stick model. But it's often perceived as beating the schools with a big stick. I suppose that is the negative side. (Bella)

Kent pointed out that 'fear' may not be 'a bad thing' in order 'bring people into line'. Kent pointed out that pressure is not necessarily negative but rather a motivation for schools to comply with Ofsted's expectations and take action to improve:

there's always been an inspector of sorts, but not on the scale that we have now. I think, when inspectorate was adopted, there used to be a sense of fear about it. And people felt frightened by it, and it did bring people into line. And I don't think that's a bad thing. (Kent)

Overall, high stakes accountability was seen by these participants as 'needed' in England, despite 'distortions' that it may create (Eduardo). Fear and wellbeing issues arising from accountability-related pressure were linked to the leadership of headteachers who are not confident about the quality of the school's provision. As discussed in Chapter 2, there are studies which make some recommendations for coping with accountability pressure and, similarly, explain that accountability pressure also improves schools (Barber, 2004; Elmore, 2006; Schleicher, 2012). Amongst the participants, two (Neil and Eduardo) argued that the current system is 'probably' based more on accountability than support (Eduardo). Thus, they pointed to the need for a 'sufficient' and 'coherent' support mechanism in the system (Neil). This point is explored further in Chapter 8.

### 6.2.4 Providing the 'right information' through Ofsted reports and school league tables as a lever of school improvement

Another benefit of Ofsted cited by some participants was related to school choice as the parents' 'right' (Bella, Eduardo) which benefits both schools and parents. Eduardo for instance saw an 'ethical dimension' in parental choice. Parents, 'as free individuals', should have 'some choice of the education of their children' and thus should be allowed to choose a school. In principle, 'that could lead to a better fit between the pupil and the school. Such differentiation might be in school curriculum or philosophy and parents can thus 'select the one that is best for their particular children'. Neil pointed to how parental choice 'makes perfect sense' in 'theory', but he also noted how it does not for 'certain geographical locations.' Neil explicitly problematised parental choice and explained this weakness as an 'area of market failure'. Although he considered calling parents 'customers' or 'users' as 'cold', his approach to parents as 'acting in the best interests of their child' (rather than in the best interests of the community) suggests an underlying individualistic approach:

> It makes perfect sense in theory, and it makes sense in certain geographical locations. It does not make sense in a town in Northamptonshire that only has one secondary school. Or rural Rutland where there's only one primary school within 10 miles. It doesn't… That's an area of market failure. And, I think, there is more market failure than has been acknowledged due to geography. So, in theory, fine and the problem is in practice… And I don't like 'customer', it is quite a 90's word. I don't like 'user' either because it's very cold. But, thinking of parents as parents who are acting in the best interests of their child. (Neil)

Bella pointed out that 'a vast majority of parents… get into a school that they want to get into', provided they consider the school's admission criteria. Kent argued that, despite a lot of 'shouting' about parental choice, there is 'a significant amount of inertia' and many parents follow 'tradition' rather than exercising choice. He shared his experience in his current school, where the 'overwhelming majority' of children come from 'within a catchment area that already exists'. In a similar vein, Bella, who is currently working as a school adjudicator, argued that 'the vast majority' of parents can send their children to their desired schools in England unless they are 'unrealistic' about their choices and the admissions criteria of the school. She noted that 'objections to admission' are 'bound to happen' because 'popular' schools are oversubscribed. Further, she argued that parental choice is 'a great idea':

> I think, it's a great idea. I think, I've always considered parents to be my customers and if parents don't want to send their children to my school, then I would be very, very worried. (Bella)

Overall, the above experts' views were aligned with the official discourse of parental choice

presented in Chapter 2 (Ofsted, 2015, 2019a), based on the importance of 'providing' parents with information when giving them a 'right' to choose a school. Both Ofsted and performance tables were deemed useful for parents as a source of information, though they also seemed to be cautious about the sole use of league tables to inform parents. Performance tables are the 'end product' (Neil) which 'provides clear information to parents and other actors within the system' (Eduardo). They, however, also referred to their limitations, for example Neil pointed out that performance tables may be difficult to understand by some parents. Similarly, Bella pointed to the issue that performance tables are 'not properly value added', as if each school were 'equivalent' in 'all social deprivation areas and all ethnic minority areas.' Thus, she argued that league tables 'don't really tell you anything much more than the area in which the school is situated'. Bella (like Kent) found it 'unfortunate' that league tables might 'influence' some parents. However, Bella, as a previous school practitioner and current *policy actor*, also believes that today's parents are 'far more savvy' about league tables and 'they will have a look at the school's Ofsted report which is very sensible' because 'we've got the internet'. As she explained:

> What we need to do is provide parents with the right information that gives them a proper parental choice rather than what they said in the local pub or the local cafe. You know, 'Oh, that's an awful school because my son did this bad.' (Bella)

Bella's argument about the importance of Ofsted providing the 'right' information' for parents resonated with Eduardo's and Neil's arguments that Ofsted can provide 'understandable' and 'trusted' information to parents about the quality of schools. Eduardo pointed to the importance of making reports 'easy' and 'understandable' to parents as a responsibility of Ofsted. Neil pointed out that Ofsted's interest is to 'stick up for parents' and serve as a 'trusted source' to parents, who need to know how good the school is.

A combination of the 'right' information and parental choice was deemed a key 'lever' (Neil) providing schools with an 'impetus' (Eduardo) to improve. Eduardo explained the importance of informed parental choice and gave an example from a country which has parental choice but does not provide information about the school and does not have national tests. He argued that, in such a system, a school's success or failure is a 'self-fulfilling prophecy' because the same schools tend to retain their established reputation:

> [a country has] school choice like in England but… doesn't have any national tests or indeed much other information about the school. So, on what basis then do parents make that choice? Well, essentially, on the reputation of the school. Umm, now, that is of course very much… That, that has problems. Firstly, it's a system which is self-fulfilling prophecy because always it is the same schools who attain the same reputation to a large

extent. Secondly, that doesn't provide much of an impetus for those schools to improve. I've got reputation, I'm therefore gifting umm most motivated parents choosing to send their children to this particular school and I can coast along on, on my reputation alone. And nothing will harm me in that respect unless something terrible happens and my kids get caught dealing drugs or something like that. So, tests are important in terms of providing information. (Eduardo)

Neil noted that parents can put pressure on schools as a 'leverage' for improvement while holding schools accountable through the information available to them:

> We [Ofsted] are giving information to parents to enable them to hold schools to account. Sometimes that takes place at the level of school choice… But sometimes, it is: 'I am a parent of a child at the school, and I have seen the Ofsted report, they're not teaching and reading in the right way. I'm going to use that as leverage to pressure the school to improve'. (Neil)

Neil's argument for holding schools accountable through parental pressure echoed Kent's experience as a school practitioner. Kent highlighted that schools need to 'maintain numbers' because it is about 'reputation':

> a lot of parents, not all but a lot of parents, gauge their school places on how our school performs. Even that's only one measure. You know, cross small measure to be fair. But it is a measure and parents take it very seriously. So, I think, there's a reputational issue which we take very seriously which we have to make sure that you know we cover. (Kent)

An important finding which emerged from interview data pertaining to questions about high stakes tests, school league tables, parental choice, and marketisation-based strategies (Appendix 2) is that these participants did not talk much about marketisation. Only Bella referred to the benefits of 'school funding [that] follows pupils' and makes schools 'bound to market their schools'. As she put it:

> As you know, school funding follows pupils. So, if you've got a plan… if you've got a published admission number of 180 and you only get 150 children into your school, then you've lost out on 30 times the amount of money that you get for a child. So, schools are bound to market their schools. They need to get the children in. And if you don't do that and if you fail and if your school numbers go down; then, you don't have the money. Therefore, you don't have the resources; therefore, it will get worse. So, I think, it's a good system… (Bella)

Competition was another notion rarely articulated by these experts; except for Eduardo, who explained that 'the competitive mechanism is seen as a motivator and a spur towards improvement' because, by 'allowing parental choice and making sure that funding follows pupils, you make schools ensure that they provide what parents want, which in most cases is good quality education'. Eduardo also noted that tests and parental choice allow schools to 'compare themselves to each other' and 'the combination of parental choice and publishing

performance tables' could have a 'motivation function in terms of getting them to improve and getting them to work on test results.'

Overall, these arguments about the importance of 'providing' parents with information when giving them a 'right' to choose a school resonated with the literature discussed in Chapter 2 (Burgess et al., 2014; Ofsted, 2015, 2019a). Both Ofsted reports and performance tables were deemed useful for parents as a source of information, though the sole use of league tables to inform parents was approached with caution. Whereas choice was suggested as limited due to the 'geographical restriction' (Neil), there was consensus that the vast majority of parents benefit from transparency and can send their children to schools 'they want' (Bella). Importantly, providing information to parents not only works for parents but also works for schools. Parents can put pressure on schools to improve and schools are thus accountable to parents. Also, publicly available information can provide an 'impetus' for schools to improve. Competition, marketisation and pupil funding were rarely mentioned by this group of participants, except for Eduardo. His interpretation of the competition led him to question competition as discussed in Chapter 8.

## 6.3 Problems arising from the official approach to school evaluation

This chapter now moves to discuss findings pertaining to problems with the official approach to school evaluation in England. As explained in Chapter 4, a general tendency amongst participants who expressed predominantly positive views was to refer to 'standards' and 'frameworks' in terms of their benefits such as creating 'clear' frameworks for objective evaluation. In contrast, participants with predominantly critical views approached 'standards', 'frameworks', and 'criteria' with caution and highlighted their negative consequences (e.g., undermining creativity and professionalism) and limitations (e.g., considering schools as if they operated under the same conditions), as well as their nature as 'rigid', 'prescriptive', 'too idealistic' and coming from 'above' (Ned). Overall, this study found that the problems arising from the official approach to school evaluation since 1988 included: 'prescriptive frameworks and externally set, rigid criteria' (Kelvin, Cooper, Ned, Torr, Orion, Nora, Dennis, Felicia, Gordon), 'problems with the concept of accountability as a driver of improvement' (Cooper, Nora, Ned, Torr, Kelvin, Orion, Gordon, Dennis), and 'marketisation as a recipe for individualism, lack of coherence and poor governance of education' (Dennis, Cooper, Kelvin, Orion, Felicia, Gordon, Nora).

### 6.3.1 Prescriptive frameworks and externally set, rigid criteria

Problems arising from standards and frameworks cited by some participants were not about having national standards, but about how and why to use national frameworks and standards (i.e., rigidly, to make judgement, as political tool). In this regard, whenever benefits were articulated, they were followed by expressions of concern and reservations pertaining to 'set' criteria (Torr). Kelvin stated that he found 'understandable' the changes in the 1988 Education Reform Act in terms of 'introducing a bit more of a national framework'. Although he found inspection frameworks 'quite useful', he also emphasised that 'it is not an exclusive yardstick by which you can judge how you are doing as a school'. In a similar vein, Cooper, as a *policy actor* who was a managing officer in a local authority tasked with introducing the reforms that come along with ERA and the establishment of Ofsted, said that schools were 'maybe… too free' pre-1988 and national standards were 'missing'. Although it is 'good that there is a national standard', it is 'not enough'. In a similar vein, Dennis, *a policy actor* who worked as an HMI Inspector before and after the establishment of Ofsted, noted that inspectorate before Ofsted 'made judgements about schools, but never made it clear what the criteria were that they were judging the schools against'. He pointed out that it is a 'good thing' to be 'open' in terms of 'what the inspectors are looking for' and Ofsted frameworks and handbooks are 'useful in one sense'. However, he immediately problematised inspection frameworks as idealistic'. As he put it:

> to be honest, Ofsted are expecting too much of schools and too much of their inspectors. So, although they are open about what they're trying to do, I think, they're probably trying to do too much. And their frameworks and their handbooks, though useful in one sense, are perhaps rather too idealistic. (Dennis)

Dennis's argument was echoed by Felicia and Ned as *school practitioners*. These participants critiqued the use of standardised national frameworks within the school context. Expecting the same from each school was 'too rigid' (Ned) and 'not fair' (Felicia). Ned worked as a headteacher and lead in a local authority and noted that 'school contexts are all very different' and Ofsted frameworks do not 'reflect sufficiently the individual contexts of the schools':

> My issue is that they evaluate against an agenda that is very, very de-contextualised… Ofsted use a very broad framework and then try and apply it in every school context. School contexts are all very different. So, some schools are working in incredibly challenging circumstances dealing with young people who come from very dysfunctional families; for example, dealing with young people who may be come from families where English is not the home language. They are obviously working in a very different way to schools where the children let's say very well supported from home, very well supported in terms of being able to speak English fluently. And for example, some schools have a number of parents who'd be able to say: 'I can get my child a tutor if I'm concerned that they're not making the progress at school that they want.' And I

fear that the framework is too rigid and doesn't reflect sufficiently the individual contexts of the schools. So, for example, they will say things like: 'If results aren't at this level; then, leadership cannot be judged to be good'. (Ned)

In a similar vein, Felicia highlighted that it is 'not fair' to expect each school to 'fit' and 'match' their 'criteria', because there might be some aspects that schools have 'very little control about'. She shared her previous work experience in a specialist school (University Technical College) where such criteria 'will never match' this school because 'that's not the kind of school it is':

> I worked for a UTC in the last couple of years and that's a University Technical College. Now, that's a specialist school, and intakes in year 10. 98 percent boys and focuses on engineering. That school got placed into 'requires improvement'… because it didn't fit the criteria that that school was never going to fit… It was morale-killing. But that's because they were using an evaluation process that compared us to schools nationally. And we were not a national school. Ofsted works on the basis that you have a nationally representative sample. What school does? Where is that school? (Felicia)

As can be seen from the points made by Felicia and Ned, prescriptive, one-size-fits-all criteria were considered 'too rigid' (Ned). Gordon observed that 'Ofsted is actually close to an admission… that they have not been treating schools in challenging areas equitably'. As he said:

> For example, a head writing to me who has 57 different languages being spoken in his school. When the inspectorate arrived and looked at the classes of English, they said: 'I'm sorry we're not prepared to take into consideration the fact that there's 57 languages being spoken'. They've either got to pass the tests or not. Now compare that with school in the North of England, where there's only one percent of the population are from ethnic minorities. Only one percent is very different than London. But the Ofsted inspector said: 'we are not allowed to take context into consideration here. We just look at the results. And the results are poor.' Ofsted is actually close to an admission in their latest reports that they have not been treating schools in challenging areas equitably. If you look at Amanda Spielman's latest paper, the third paper on the curriculum, in the middle of that, she makes that admission that they may not be treating schools in challenging areas equitably. Well, that's quite a big admission. (Gordon)

The decontextualised national standards were also problematised by Orion, who noted the importance of the 'consideration of the political and financial environment in the role that these play in school outcomes' and argued that this is 'the principal thing, that's missing' in England. This point will be further elaborated in Chapter 7, Section 7.4.3.

Another reservation expressed by these participants was about the negative influence of criteria and frameworks on schoolwork. The participants pointed out that improvement can occur through set criteria and frameworks, but this limits improvement to these 'set, specific criteria' (Torr) by decreasing the 'creativity' of teachers and schools (Cooper, Kelvin). Dennis, *a policy actor* who worked as an inspector, expressed his reservation about the influence of frameworks

on school practice. Dennis shared 'a famous saying amongst us, inspectors' which is that: 'people do what you inspect, not what you expect'. He explained that 'rightly or wrongly', Ofsted creates an agenda on school evaluation and 'there's not been too many aspects of school [self] evaluation which have not been Ofsted related'. This might be 'useful' but 'narrow', because 'schools have been often anxious simply to do what Ofsted expects them to do' rather than 'thinking out their own priorities and their own judgments'. In a similar vein, Nora, a *policy influencer* who worked as a senior research officer for an 'awarding body' ('exam board'), said that she began to notice a significant change from the 1990s with teachers becoming 'anxious' about linking their work to the criteria for grading schools. As she put it:

> Somewhere I began to notice that what was happening in schools in the 1990s were very, very different to what it had happened earlier on and even the way that teachers behaved. I noticed that teachers were behaving very differently... So, they were becoming quite anxious. Whenever you did work with them, they were always asking questions like, 'How, how can I make sure that what you are giving is linked to the grade criteria? How can I link this in the school accountability measures?' And, at that time, I had an interest, but no knowledge of assessments. So, most of the time, I was saying 'I've got no idea [laugh]. That's your job'. But I was then told that it was very strange that it was their job. That's something to me that didn't seem very much about education, was preoccupying them to great deal. (Nora)

In a similar vein, Torr found that set criteria 'shift the focus of the school leaders' to improving 'some things', but this neglects 'other things that they could have done'. Torr also found 'set' criteria problematic because having a 'rigid framework with more criteria' is a sign of lack trust in schools and teachers. As he put it, criteria:

> clearly have an impact on the way that schools have been managed, how schools manage their workforce and how they manage the students. All of which make a difference to the quality of education. So, as a result of the criteria that they're using to check compliance, schools have moved and improved for those specific criteria. But because they've set very specific criteria, schools have chosen to *just* [emphasised] meet those criteria rather than necessarily thinking about other things that might have also had an impact on education. So, it shifted the focus of the school leaders into improving some things, which has been good. But the bad side is obviously that that leaves other things that they could have done. (Torr)

Torr's point above resonated with Dennis's, Kelvin's, Ned's, Felicia's, and Cooper's views. For instance, Kelvin, as a *school practitioner*, shared how he used the framework in a 'very challenging school'. He assigned roles in the school according to the Ofsted framework, 'using it as a yardstick'. Interestingly, he did not make it clear whether this helped to improve the school, but he noted that the framework helped the school in terms of 'knowing what was seen as a good' way of 'doing stuff':

> I think, the framework is quite useful but not an exclusive yardstick by which you can judge how you are doing as a school... almost by accident, in my second headship, I

was able eventually to appoint to my team people with different roles and actually looking at their strengths. I matched them to a section of the Ofsted framework. So, when the Ofsted framework was something like teaching, something like behaviour; then I had someone in charge to teaching and I had someone in charge of behaviour and someone in charge of progress. So, when it was kind of divided up like that, my, my, my, it helps in terms of just pulling everything together, providing a degree of focus, using it as a yardstick. And, in terms of on the shop floor, it was with colleagues and then colleagues had something they could go to in terms of knowing what was seen as a good, way of doing stuff. That's the benefit, I suppose. (Kelvin)

Kelvin's reference to 'I suppose' suggests a reservation about the benefits of Ofsted inspection frameworks. This reservation was confirmed by his comment that he found the framework 'very prescriptive' because it presents a 'very, very rigid view of what a school should be like'. He also emphasised that 'external pressure' on schools 'was linked to people losing jobs' and argued that inspection frameworks, together with external pressure, are a 'complete opposite' of 'mastery, autonomy and purpose', as well as 'creativity':

> The negative is it's, it's very prescriptive. I think, one of the upshots of a lot of developments after 1988s was a reduction in creativity. And also, the whole external pressure which was linked to people losing jobs. So, I was sitting in a meeting in this city with headteachers and then a couple of months later we had another meeting, and someone had disappeared and then another one had disappeared… headteachers were losing their jobs as a direct result of this and the pressures associated with it either through illness or because they were removed or because they got bad Ofsted and then they were removed, or they left before they have to. That's a sick system, it's not a healthy system... it is complete opposite of autonomy, mastery and purpose… (Kelvin)

In a similar vein, Cooper pointed to the link between 'fear' and 'courage to step outside' of the frameworks. He said that, as a result of the changes after the 1990s, 'what we have now' is 'fear' and people who 'tend not to do anything other than that which is going to contribute to the next Ofsted inspection.' Thus schools 'prioritise those things which they are inspected on and ignoring the things which they are not inspected on', in order to make sure that 'what the Ofsted inspectors look at when the Ofsted inspectors come is good' and 'do those things which will result in a successful Ofsted inspection next time'. Therefore, schools 'can't afford to put time into broader education'; that in turn restricts school practices such as preparing children for living in a society.

The participants also referred to 'too many changes' in inspection frameworks that created huge additional workload for schools and confused both inspectors and schools. For instance, Dennis pointed out that schools since 1992 have been 'too often' trying to 'guess what Ofsted wants' or 'is expecting'. Cooper's and Dennis's arguments about excessive workload arising from 'too many changes' resonated with Kelvin's experience, who is a school practitioner and

received eleven inspection visits in total in his 12 years headship. Nine out of these eleven inspections were in his last five years (six of them from Ofsted):

> From about 2012 onwards, the framework kept changing much more frequently. To this extent, there was one year where there was a different framework every term. For schools like ours, which would do an Ofsted inspection at any time, that meant that you had to spend a huge time, familiarising yourself with the framework, which was a huge document, the evaluation schedule, including safeguarding but quite different documents. Just to make sure that you have been in a job at the end of that session. (Kelvin)

Another concern expressed by these experts was about the content of inspection frameworks which has been used in 'reality' (Kelvin) to promote the 'government agenda'. Gordon, for example, questioned the independence of Ofsted as follows:

> Ofsted claims to be independent, but it is an arm of government… in practice, every time the government asks for a change say in education like 'We want the schools to promote British values of democracy, fairness and against radicalisation', immediately, Ofsted puts that into its framework for inspection. (Gordon)

In a similar vein, Kelvin described Ofsted as 'a tool' used by policy makers to 'enforce a particular agenda'. He cited an earlier government agenda to make every school an academy by 2020 and explained how Ofsted grades were used to force schools to convert into an academy. He said, 'there was a spate of bad inspections in this 2012 to 2013 which led to a lot of those schools that were graded as 'inadequate'. Joining the academy system which was of course government policy'. Using standards and inspection frameworks as tools for promoting political agendas is further elaborated in Chapter 8, Section 8.2.1.

Overall, participants had reservations about criteria, standards, and frameworks because of the problems with their use and the reasons for their use. Even though transparency and openness about the criteria that schools will be 'judged' against are a 'good thing' (Dennis), inspection frameworks are not 'exclusive yardsticks' to 'judge' how schools are doing (Kelvin). Considering the times pre-1988, having national frameworks and standards is 'understandable' (Kelvin) but not 'good enough' (Cooper). Ofsted frameworks have been 'very prescriptive' (Kelvin) and 'too rigid' (Ned, Torr), presenting a 'very rigid view of what school should be like' (Kelvin). Frameworks and criteria shifted the focus of schools towards the expected, improving 'some things', but also limiting improvement to these 'specific criteria' (Torr). Importantly, external frameworks may help schools to know 'what was seen as a good' (Kelvin, Cooper, Kelvin) but they may also undermine the mastery of professionals. Because of the 'compliance check' (Torr) of Ofsted, schools prioritise and do what Ofsted inspect rather than working on 'what they think is necessarily right for their children' such as preparing students

for living in a society (Dennis). This narrows education practices and decreases creativity. The system which relies on 'fear' discourages people from 'stepping outside' the framework. As discussed in Chapter 2, there are studies which also suggest how standards limit the creativity of schools, but further suggest that these schools became alike organisations ('mimetic isomorphism') as they internationalise the standards over time (Ehren, 2016).

Further, a 'rigid framework' means a lack of trust for school professionals who are the 'expert' (Torr). Ofsted frameworks are 'too idealistic' and expect 'too much' from schools (Dennis). Expecting from every school the 'same' by 'fit[ting]' set criteria is not fair. Schools are different and can, at times, have 'very little control' (Felicia) over their situation. Schools, particularly those which are subject to more external evaluation (i.e., graded as 'inadequate'), spend 'huge' amount of time to 'familiarise' themselves with 'frequently changing frameworks' (Kelvin). Due to the power relations between the Government and Ofsted, the framework reflects 'government agenda' (Kelvin) and this, in turn, makes Ofsted 'a tool' to 'enforce' (Gordon) governmental agendas. The credibility, competency, and background of inspectors, for example whether they had experience of teaching, were also pointed out as important elements in designing and implementing the frameworks. It was claimed that inspectors' professionalism was 'constrained' by 'compliance culture' (Torr). These issues will be discussed further in Chapter 8, Section 8.2.1.

### 6.3.2 Problems with the concept of accountability as a key driver of school improvement

Whilst accountability was considered to be an important element in a public education system, some participants questioned the 'design' of the official accountability system, pointing to its negative impact on schools. The 'design' of the accountability system refers here to accountability measures and mechanisms of its implementation. The core problems with the official concept of accountability discussed included: externally set accountability based on parental choice, league tables and inspections that turned out to be 'counterproductive' (1); outcome-based accountability measures that created 'perverse' incentives for schools (2); accountability to Ofsted and government rather than to children, parents, and communities (3), and the punitive, flawed mechanisms of its implementation (4).

The first problem cited by participants was to do with the concept of accountability as a key driver of improvement, based on parental choice, league tables and school inspections. As such, the established accountability system has been assessed as unsuccessful in driving school improvement. Although, in exceptional individual cases, accountability 'may' have

contributed to improvement (Cooper), the 'aggregate' impact of accountability on school improvement was 'negative' (Orion) and 'counterproductive' (Kelvin), due to the narrowing of the curriculum; taking up much of teachers' time and energy; diverting schools from their own, specific, context-based needs, and causing excessive levels of stress. For instance, in Kelvin's experience, the official accountability system was 'counterproductive', stressful and described as 'being on a knife edge'. Similarly, Nora observed that during her frequent school visits in the 1990s she found it 'very strange' that the role of educators was displaced with activities such as finding evidence of the school meeting accountability demands. In a similar vein, Orion pointed to the 'money and energy' put into 'displacement' activities:

> As a driver of school improvement overall, it's pretty unsuccessful. So, does accountability drive up standards overall? No. No, in fact it probably has slightly the reverse impact. It's because it takes money and energy and puts it into a displacement activity… But in individual examples, it can make a difference. So, there's difference between the aggregate and the case. So, there are some instances where accountability has indeed meant that the institution has to improve and has been played a fairly vital role in establishing the basis for improvement. But that's quite different from what the aggregate impact is. Aggregate impact of accountability has been negative… I don't think it's played a significant role in improving the quality of education rather the reverse. (Orion)

When I asked Cooper about 'his views about accountability as a driver of school improvement', he said that 'it's almost the wrong question in my view'. My question was problematic for him, because he was against 'the idea of producing school improvement' through 'overtly public' external accountability. He emphasised that 'in the name of accountability' England introduced 'inspection, performance tables and alongside that is parental choices', so that 'we have this external accountability through the publication of inspection reports and appalling league tables.' However, for him, teachers and schools are 'already self-improving' and 'they do not have to be driven'. 'External accountability' as a key driver of improvement 'presumes that professionals don't care' and introduces 'the mechanisms of the factory into a professional area'. This system does not consider the 'local' and internal 'school values.' As Cooper put it:

> Of course, any publicly funded system should be accountable to the public. But the question is 'How?' 'Through what mechanism should it be accountable?' It is public money. So, what we have introduced in the name of accountability is inspection, performance tables and, alongside that, is parental choices which it isn't. It isn't choice. There's no choice… So, if I understand your question, you're asking me from the perspective perhaps internationally of whether having a system which is overtly publicly accountable as ours appears to be by having inspections and league tables etc.; what purpose of that purpose is? Well, it ought to be improvement. And I've no doubt that being accountable has an effect on schools. Schools are more alert more aware than they were; about the need to continuously improve. But I would only say, if it's the only question, is it the right question? Because it presumes that professionals don't care. It is introducing the mechanisms of the factory into a professional area. My view is that

ninety nine percent of people who teach want to do a good job for children. So, they're already self-improving. They already think like that. They don't have to be driven. The accountability system is an external driver, and it isn't the best mechanism for improvement. OK? So, accountability, it's one of those things that you can't say: 'it's not a good idea'. It's a good idea. But it's not the idea to produce school improvement. (Cooper)

The second problem pertaining to accountability was that outcomes-based accountability measures created 'perverse' incentives for schools, teachers and students (Ned, Felicia, Kelvin). Since these measures were mainly based on student outcomes, accountability 'boiled down to looking at test scores' (Dennis, Gordon). Excessive emphasis on examination results, was seen as central to 'all the problems that we have in England now', i.e., the narrowing of the curriculum, teaching to the test and 'gaming the system' (Gordon). Dennis argued that 'in England… tests have become a substitute for the curriculum in too many cases' and thus 'the examination system has certainly determined that the curriculum of many-many schools too'. High stakes testing resulted in teachers and schools 'teaching to the test rather than teaching what they think is appropriate' (Dennis).

Outcomes-based accountability measures have thus created 'perverse' incentives for schools, teachers, and students (Ned, Felicia, Kelvin). Ned explained that when 'people fixate on certain performance criteria', they 'will do everything they can to get up the league table'. Felicia claimed that 'as long as accountability exists in terms of exam results, you'll never get rid of the cheating'. Ned further pointed to examples of school leaders 'acting not in the interests of the child but in the interests of the organisation in order that the organisation can move up to the league tables', i.e., exclusion of students. Whilst performance measures may be helpful in business settings, education is not a 'business':

> We don't deal with product. We deal with young people's lives and those lives are unique, and individual and they change. We're dealing with human beings. And it is not reasonable to constantly be measuring achievement and using the very blunt tool of 'What did they gain in English? How much progress did they make in maths?' and that way making a judgment… it's all too easy to say that education is like a business and the accountability measures therefore are like the balance sheet, the profit and loss. And whilst I don't think for a moment that schools should be unaccountable, I don't believe that the easy to measure accountability processes that we have in place are the right accountability processes. (Ned)

Felicia argued that tests and test scores are 'only one measure' of a school's performance, but the current outcome-based accountability system overlooks 'some phenomenal things' that some schools provide for their students. A consequence of outcomes-based accountability is the 'narrow' perspectives on school improvement, 'drilled into us' (Felicia) and undermining

'the whole life of the school' (Dennis). For Dennis, 'school improvement' is 'far more than just simply test outcomes' and whether accountability improves the quality of education depends on its underlying values. As he put it:

> Certainly, the government would see that its accountability system has driven school improvement. Because the government of this country believes [laughed] that you can measure the quality and the standards in the schools through tests. I personally believe, you can't do that. I personally believe that accountability through tests gives you some information about a school or schools but is not an adequate mechanism for getting… assessment and evaluation of what a school or what schools can do. (Dennis)

Felicia paid attention to international comparisons and questioned why results in exams such as PISA are taken to mean that a highly scoring education system is better. For her, school improvement was also based on improving student wellbeing. As she put it:

> I think, Britain looks internationally all the time, and we seem to compare ourselves to other countries all the time, particularly Singapore… we keep looking and saying: 'Well, we're not as good as them'. But we keep just looking at exam results. And I don't think that's what we need to be looking at. You know, like PISA results. So, is that really what we want to emulate? And if we are just focused on higher and higher and higher and higher exam results, what for? Why is that better? Why does that mean that our education system is better? Particularly, if that's combined with growing mental health issues, growing anxiety amongst our young people we're seeing already. They are struggling to cope. Just before this interview, I've just had two students in tears. They've got a finance exam now. I've just tried to get into the exam room because they're frightened, and they're stressed. Is that what it's for? (Felicia)

In a similar vein, Ned noted the negative impact of high stakes tests on students, both their learning and their wellbeing:

> the process we currently have is provoking mental health issues, is causing young people just to fixate on short term memory, is insisting that they are put through extreme pressure at the age of 16 and younger as they do mock exams, and they do tests, and they do a whole barrage of different examinations. (Ned)

Felicia found it 'quite demoralising' to measure teacher effectiveness in England 'by the grades that teacher gets' because this misses lots of contribution that they are making to the students. Her teaching experience in New Zealand 'surprised' her because there were no 'target grades' for students. This made her reflect on: 'What a lesson was, how to engage, in what you're teaching, why you're teaching it, what for?' Teacher evaluation was 'more child centred' and was more like a dialogue about the lessons, about the students and their progress and engagement. The inspectorate was more interested in 'the teaching and learning and what's going on in the classroom and the leadership of the school' and she was not asked for 'exam results', or 'predictions' of grades. As she explained:

> I did seven years in New Zealand, teaching in all boys school over there. And I was really surprised. Because I turned up and saying: 'Where are the target grades? Where

are the... How do you know if I've done a good job?' And they were: 'Well, we don't have them'. And I nearly fell off my chair. So, I said: 'What do you mean you don't have target grades, they don't exist? How do you know if the students have done a good job?' They said: 'All they pass'. And it was like, it was like [laughing] a revelation to me. Because I had only in that point taught in the UK and I thought that was education. I didn't have any broader context. So, they still have an Ofsted equivalent. But it's not about exam results. It's much more about the teaching and learning and what's going on in the classroom and the leadership of the school. So, that was quite interesting in the school evaluation process felt much more supportive than it has ever felt here in the UK… When I turned up in my New Zealand classroom, I think, that experience taught me more about teaching than anything else. Because all the assumptions I've made what teaching was… Some of those assumptions still existed. But lots of them didn't. And I had to rethink: 'What a lesson was, how to engage, in what you're teaching, why you're teaching it, what for?' (Felicia)

For Orion and Felicia, Ofsted's 2019 Framework's promise to be looking at the quality of education and the curriculum was a 'welcome change'. However, there were some reservations about its implementation. For instance, Orion pointed to the importance of how the notion of 'quality of education' is operationalised in terms of the quality of educational relationships:

Focusing on the quality of education as opposed to the performance indicators is doubtless well-intentioned. But until we see what that actually looks like, it's difficult to know how Ofsted and school governors will be able to do that unless there are significant developments in the way schools are inspected and how to account for themselves.

But if, by 'the quality of education', there is going to be a focus on the quality of educational relationships and the quality of educational resources and also focusing on behaviour, exclusion, inclusiveness and things of that kind; then I think, that is also an important development. As the notion of the quality of education gets operationalised, we see what it looks like in practice. What does Ofsted do? What is the school asked to produce? and how are school governors and others responsible for the quality of education involved in there? Once we can see that; then, I will be pleased if it is done in the experimental and provisional way learning about its consequences and modifying it as we go along. And I will be extremely concerned if it is written in stone and unchangeable.

I quite like some of these additions that Ofsted has been talking about. So, I think, the richness of the curriculum might be an important criterion and that may vary from school to school, well, at least from school environment to school environment. You know, so, what you're able to do in a small rural primary school on the edge of Loch Lomond is not what you can do in a large primary school in the centre of London. I think wellbeing indicators for children would be important, mental health indicators, exclusions, expulsions, those sorts of things. But I'm not sure that it would create less pressure for the school. But they might make schools a tend to different things. You know, I really worry about children being excluded from education because they're not gonna perform well, and so they get excluded. (Orion)

The third problem noted by the experts pertained to the tendency to hold schools accountable to Ofsted inspectors and the government rather than to children, parents, and communities.

Ned, for instance, explained that he does not 'have a problem with accountability' but has a problem with external accountability measures 'from above':

> I don't have a problem with accountability. It's about what form does that accountability takes. So, I would say to you: 'I am accountable for the children and the parents who I serve'. So, if a child is unhappy, they come and see me. If a parent is unhappy, they come and see me. And they tell me what the problem is, and together we try and work through to get a solution. But the accountability measures come above that. They come from the likes of Ofsted, come through the examination system, come from Multi Academy Trusts, come from local authorities. (Ned)

Cooper, Felicia, Nora, Gordon, and Torr viewed accountability measures that come from 'above' as a political tool. For example, Nora and Cooper pointed out that accountability measures with exam results and league tables appeal to middle class voters and, in addition, enable the Government to steer the system in particular directions. However, those academic outcomes-based measures are not helping schools, teachers, and students. Torr described accountability as an 'aggressive hard-nosed approach to manage people', without considering school's own 'expertise' and 'needs.' Kelvin and Gordon argued that the whole system of evaluation has been established on market values, with Ofsted as 'a tool' used by policy makers to 'enforce a particular agenda'. Kelvin illustrated this as follows:

> going back to when Ofsted was introduced in the nineties, it was introduced as part of the Conservative government's attempt to make education more like a market. So, it came in, at the same time, there were league tables, and it was an attempt to provide information for the customer or parents or consumers as to which school you should choose… You could step back from that and say educationally 'Well, it's quite right that parents should have a clear and objective view about what schools are and therefore having some kind of national system of clear criteria for what constitute a good school or a satisfactory school etc. It is quite helpful…' But… it has become a tool that is used by various policy makers one way or another to force a particular agenda. And, particularly in this part of the world, there was a spate of bad inspections in 2012 to 2013 which led to a lot of those schools that were graded as 'inadequate'. Joining the academy system which was of course government policy. (Kelvin)

As illustrated by Kelvin's point above, most participants in this group problematised political interference in education, linking it mostly to marketisation. Some participants (Orion, Gordon, Dennis) even suggested depoliticising education.

Using accountability as a political tool shifted accountability away from children, parents, and communities to whom schools are accountable first and foremost. As Kelvin pointed out, 'school accountability, in this country, is largely governed by Ofsted', a situation that he assessed as 'wrong'. For Cooper, there should be a clear distinction between accountability and inspection: 'it is fine to have a national inspection provided that that's not the accountability system', because 'inspection is not accountability'. Although Cooper noted that

'schools have to be accountable', 'pernicious' problems arise when school accountability for broader educational purposes is conflated with schools narrowly accounting to Ofsted:

> Schools have to be accountable. It is public money of course there should be accountability. But it should be depending on what you value… I think having a national inspection system is fine, provided that that's not the accountability system. That isn't accountability… Schools do a lot of work preparing for the next Ofsted inspection which is very bureaucratic and requires a lot of information, a lot of evidence. It diverts attention from what schools should be doing and what teachers should be doing because of the amount of work involved. It results in people in schools prioritising those things which they are inspected on and ignoring the things which they are not inspected on. So, it means that the curriculum is or tends to be narrow. People do not have the courage to step outside the framework of the national curriculum. They don't do things that schools used to do like prepare children for society. They tend to do those things which will result in a successful Ofsted inspection next time. Because that's the accountability system. So, it's pernicious. It is wrongheaded. It is narrowing. It is not education... So, it is inevitable in a system where you have the level of importance attached to the inspection that we have in our system, that is not necessarily helping broader education, it narrows what schools do and it creates a system of fear and teaching to the test. (Cooper)

The fourth problem pertaining to accountability was the punitive, flawed mechanisms of its implementation that include pressure and sanctions such as losing a job or school closures. The 'risk of accountability is too strong' (Felicia) and does not allow schools to justify their decision and choose more 'creative' (Nora) responses to their changing environment. Nora and Cooper argued that the exodus of teachers from the profession was a knock-on effect of accountability pressure. Gordon claimed that the English system is 'extreme' because it is 'toughest' and 'hardest' on schools, teachers and students and the high stakes around outcomes create 'more harm than good':

> the English system is rather extreme and different from other countries. If you take the league tables for a start and when you base everything on exam results, you then get schools finding ways of 'gaming the system'. So, as you'll know, there's been a lot of this in the press recently. Some schools off-roll pupils. By which I mean, they push kids out of the school, exclude them for longer and longer periods, or they make sure they're not in the school when Ofsted is inspecting. There are all sorts of way of boosting your scores by sharp practice, by unprofessional practices. But the schools will say: 'Yes, but we are driven to this because it's so high stakes'. The senior staff are at risk of losing their jobs. Schools are closed. This is why, I consider England to be extreme. There's no other system which behaves in this very, very harsh way. (Gordon)

In a similar vein, Torr has 'a very dim view' of how accountability can be viewed and implemented in rigid ways to punish professionals 'when something isn't as expected':

> I have a very dim view of the way that some people view accountability. I don't like other people's way of looking at it. They say: 'You are accountable for this'. The implication of that is: 'They are going to hold you to account for that thing.' And that means: 'When something isn't as expected, you [will] account for what that is'. But that isn't actually the process that happens. What they say is: 'You're accountable for this and

if it doesn't work, if you don't do what you're supposed to have done, we will sack you.' But they didn't used to do that. Because they weren't allowed to. But they do that now. That's not me being given an opportunity as a classroom teacher to say: 'Well, actually, it didn't work as the way you wanted it to. Because of all of these other reasons.' It just says: 'I've got to do this regardless of whatever the conditions are'.

There's been an increase in the use of the word accountability in the last ten years, maybe the last five years. It's an aggressive hard-nosed approach to managing people. And it goes along with what I said earlier on. It's not we trust you to do your job. It's the opposite of that. You are accountable to me for doing this thing rather than I trust you to go and do this thing… in an ideal world, you employ someone that you trust to do a job and then you let them do it. We are not making things on a production line. We are experts in teaching. So, trust us to teach. (Torr)

A related problem with how accountability was implemented reported by Orion pertained to holding the school to account for outcomes that are also impacted by the wider policy environment. For instance, large numbers of students do not have access to quality education because their schools are 'under-resourced'. Given under-resourcing, putting the blame for failure solely on schools is 'wrong'. As Orion put it:

I think that it is deeply unfair that schools get held to account for performance when 'the resourcing of education' plays a significant part in the results that schools are able to achieve. And I don't just mean the money that the school itself gets. But we know, as a matter of fact that significant investment in early childhood education and preschool interventions has a lasting impact on the wellbeing including the educational performance of young people. So, what governments do in preschool and post-natal care and the like has an impact on what schools are able to do later on. And so, you know, schools do not bear the full responsibility for underperformance.

I do think it's important that we have the capacity to identify failing institutions. What I think is wrong is that we can't apportion blame correctly. So, it's the school that gets blamed. And then, we are, you know, parachuting new headteachers and we find them in a multi academy trust or whatever we put it on special measures and so on and so forth. And the political and contextual variables that make a contribution to the success or failure of a school are not part of this school report. They're not part of the story that's told about the school. That's worrying because it's a partial story. (Orion)

As can be seen above, Orion was also critical of forced academisation when schools failed to meet Ofsted's expectations. And yet, some academies also struggled, for example Kelvin took early retirement as a result of the pressure he experienced as headteacher of a newly set up academy (see also Chapter 5, Section 5.5). Government expectations of 'rapid improvement' led to a proliferation of inspections in Kelvin's final year in education:

I've had a lot of accountability. I've been responsible for accountability. I don't think you'll find many people who had more… The second school I went to has a lot of problems, it was an academy, it has just been set up as an academy and there was a lot of government expectation that being an academy would lead to rapid improvement. As a result of that, as well as facing Ofsted inspections, we also had inspections from the Department for Education. They were worse than the Ofsted inspection because there

was no set script. And then towards the end of my time in headship, there were a further two types of inspection added to the accountability process. There were inspections by the Regional Schools Commissioner, and this was a post as only just come into being in the last few years. It's a post which has enormous power in this country. Most of the general public won't have a clue that these people exist. But we had inspections from them. And then finally because of the way that schools have moved to become academies, academies have grouped together in Trusts, in my final few years of headship, we had Multi Academy Trust inspections. So, the basis of trying to find out what it's like to go through the accountability process, I've gone through eight Ofsted inspections as well as DFE inspections, MAT inspections and RSC inspections…

I think, there should be one system of accountability, there are at least four currently that I've undergone as a headteacher. So, it, most of these questions about Ofsted; but DfE, RSC, MAT, all these different. Most of those other three were worse than Ofsted. Everyone sort of thinks that Ofsted is kind of so bad. But actually, there are four and of those, the DfE inspection is the worst. (Kelvin)

Overall, accountability was seen by these participants as an important element in education, which potentially contributed to improvement in some schools. However, its aggregate impact has been assessed as 'counterproductive' (Orion), contrary to government stated intention. This was mainly due to how accountability was designed and enacted, leading to four core problems highlighted by experts with predominantly critical views. The experts problematised the conception of accountability as a means of 'external' mechanisms of parental choice, league tables an Ofsted inspection with an 'assumption' that accountability improves is problematic, based on four insights. Firstly, accountability is not about improvement. Secondly, it is problematic to assume that accountability and Ofsted inspection are the same because accountability is not inspection. The issue here is that schools' accountability to Ofsted and the Government rather than to children, parents, and communities is a flawed assumption. Thirdly, outcomes-based accountability which concentrates mainly on quantitative data of students' results (i.e., attendance, test results, exclusion) is also problematic because this narrows education for broader purposes, and creates 'perverse' incentives for schools, i.e., gaming. Finally, the punitive, flawed mechanisms of its implementation are problematic. The reconfiguration of the 'design' of accountability and mechanisms of its implementation suggested by these experts are discussed in Chapter 7.

### 6.3.3 Marketisation as a 'false God' and 'recipe' for individualism, lack of coherence and 'poor' governance of education

In addition to problematising accountability, these participants also explicitly problematised marketisation-based strategies presented by successive governments as a driver of improvement. As indicated in the title of this section, there were two main strands in their

critique of marketisation strategies. The first one was the flawed rationale of marketisation as a driver of school improvement, expressed by Kelvin's metaphor of marketisation as a 'false God that we all bowed down to'. The second was captured by Dennis, who referred to marketisation as a 'recipe' for individualism, lack of coherence and 'poor' governance of education. Importantly, the key problem highlighted by several participants was the conception of a 'single' school improvement and the 'idea' of improvement through market forces that revolve around parental choice. Whereas the participants critiqued the idea of improvement through parental choice, what they saw as 'powerful' (Felicia) for improvement was the 'community support' (Ned) and 'engaging parent voice in the view of the school' (Felicia). Owning improvement should belong to professionals themselves, rather than to external market forces, with responsibility for improvement given to each school within a system where schools work together, not against each other (a point discussed further in Chapter 7).

The participants' views converged on a strong critique of marketisation strategies in England in a unanimous conclusion that marketisation did not improve the quality of education. Orion for instance, as a *policy influencer* and expert in the evaluation field, said, 'I could be neutral about the policy' but 'I can't see the evidence that marketisation has improved the quality of education'. Specifically, there were no evaluation studies on the impact of marketisation-driven changes conducted 'in a way which yield decent evaluative data'. Without such data, marketising education amounts to a thirty-year long experiment: 'given that the politicians have continued to complain about the quality of education, I assume that they are not convinced by their own experiment' (Orion). In a similar vein, Kelvin and Cooper also pointed to the lack of 'evidence' that marketisation improved the quality of education and, in addition, noted its ideological roots. They attributed such strategies to the 'right-wing competitive philosophy' (Kelvin, Cooper, Torr) from a Conservative Government in the early 1990s, adding that 'a mistake… made was that when Labour was in power from 1997… they didn't really change that' (Cooper). Cooper emphasised that marketisation 'doesn't work for those who are relatively disadvantaged, and those children end up in slightly less-good schools.'

The 'philosophy' behind marketisation and in particular the rationale behind 'competition' and its business-orientation were assessed as problematic when applied in educational contexts. Orion for instance pointed to an 'unarticulated theory of change' which suggests that since competition works for widget production, it will work for schools too. As he put it:

> We have an unarticulated theory of change. The idea here is that: 'if competition works for widget production and the competition means that you get better widgets; it will

work for schools. Competition will mean you get better schools.' But schools are not widgets (one) and (two) we don't have any evidence. There have been no randomised controlled trials. There have been no systematic reviews of 'what the impact of these changes has been' in a way which yields decent evaluative data. But we've been doing it a long time… Why would competition work for every child? I can't see it. (Orion)

Orion's views were echoed by other participants. Kelvin said that 'the premise is that competition is good… it drives up improvement' but 'businesses [also] go out of business.' According to Kelvin, treating schools as businesses 'doesn't help anyone… doesn't help the country… does not help the nation'. In a similar vein, Cooper questioned whether it would be allowed for one hospital to be better than another hospital, he did not find this acceptable in education either. As he put it:

> If that was hospitals, we wouldn't allow it. Why do we allow it with schools? So, whatever's happening with the health systems in terms of being improved, it's certainly not seen as acceptable that one hospital is better than another hospital, is it? This is just not acceptable. So, it's not acceptable one school is better than another school. It only to be as good as they can be. (Cooper)

The 'cost' of competition in public education was highlighted by Kelvin. Gordon pointed to government policies which intended to support collaboration between schools, particularly after 2010, and saw it as a 'paradox' because it is 'almost impossible for schools to collaborate and compete at the same time'. He gave an example of how two 'very good' comprehensive schools served the same city but competed with each other over Ofsted grades:

> The competition is very serious in an area where I am just now. There are two very good comprehensive schools and they both compete for numbers. And when one school gets an outstanding and the other only gets a good; parents move towards the outstanding school. (Gordon)

Kelvin, as a school practitioner, shared his personal experience of how competition created an 'aggressive war' between schools in his local area due to the strategy of a school 'to entice' pupils to go to that school. I am citing his account in full because it provides a highly evocative account of the 'aggressive edge' of competition:

> it's the time for open evening, so schools will have an evening and parents come in and have a look around. And one of the local schools published our exam results, our examinations plus exam results of two other schools… Now our exam results last year they went down, we are a small year group within that small group there are about 10 boys who were bright but underperforming and they were the most unmotivated… That's taken that figure, that year I [was] put in the local newspaper in an advert with two others... and they put their own figures. Well, they have not said about their own figures. And I'm choosing my words carefully here, it's how many of those students in that percentage that they gave started the year and how many they had lost over the course of that year. And I think that's, that's I would just leave that hanging in there… That is horrific, because that is basically… umm, it's almost industrial espionage. It is a gross misrepresentation, highly aggressive. When I first came to this city there was an

129

agreement that none of us would publish anything about any of the school's results and figures and that just being flagrantly disregarded. So, yes, that's that' aggressive, aggressive war. (Kelvin)

Felicia also problematised competition because the 'playing field' should be 'level', referring to differing contextual situation of schools and the schools' need for 'support' rather than play the game of winners and losers (see also Chapter 7). Her school had 'students who are both eligible for pupil premium and have a special educational need'. Expecting disadvantaged students to 'make the same progress as everybody… isn't helping' without the right level of support.

Kelvin also pointed to the time schools spent 'to promote and market themselves to attract more students' instead of 'worrying about academic matters.' Nora argued that schools 'waste' their time to attract more students to enrol by creating 'a really attractive' school website:

> people are having to waste time creating a really attractive school website or information for documenting, trying to enrol more pupils there. Because of course that's where your money is coming from. Whereas actually all of that time and effort could be put into producing amazing lessons and fantastic resources and in a brilliant school community. (Nora)

Felicia, Torr and Kelvin, who are all *school practitioners*, argued that parental choice creates 'not good' but 'popular' schools, and those schools turn into 'de facto selective schools', because they choose their 'intake' (Kelvin) and, consequently, those popular schools 'get better and better' (Torr). The issue of schools 'creaming off the best children' (Kelvin) was echoed by Felicia, who noted that popular schools become oversubscribed and then 'it just becomes a cycle that schools can't escape' and, in effect, these popular schools 'can't fail'. As Felicia also explained, the remaining schools that are unable to be selective 'get worse', because they have to 'mop up' everybody else. Torr's experience echoed these views:

> The school I work in has a disproportionately large middle-class cohort, partly because a lot of parents outside the traditional catchment area choose to send their children there, because they think it's a good school. And consequently, it is a good school because they send lots of nice middleclass children. So, it's a self-fulfilling prophecy. You know, if they send their kids there, yes, it will continue to be a nice school. And, you know, I know people who've moved so to make sure that the children are close enough to get into oversubscribed school. If you can afford to do it; then, you do. And so, the rich kids get richer. So, it enhances inequality. (Torr)

Applying the market idea of consumer choice in education was for these participants a 'misconception' about the mechanisms through which quality is achieved 'in a system of public education' (Cooper). This mechanism works for some schools at the expense of other schools and does not improve the system 'as a whole'. The 'fundamental mistake' here is an assumption

130

that 'schools are separate in a system of public education'. Cooper described this as a system where 'success breeds success' but does not consider the failing schools:

> it is fundamental mistake of thinking that schools are separate, but they're not... in a system of public education. And so, it is judgmental and relies on the idea of parental choice driving the quality. And the mistake that it makes is that in a public system we shouldn't be interested in school improvement for a school. We should be interested in school improvement for all schools. That's fundamentally different. So, I think, it's misconceived. I think, it's doesn't work, hasn't worked. We've had it now for 30 years and it hasn't delivered what we were told it would and which people like me [in my role] never believed it would deliver. For the quality of public education as a whole; that's the only thing we should be interested in, not of some schools at the expense of other school, all schools.
>
> all schools are connected by the children; they are the same children. So, the idea is... Because our system or mechanism is a fact... is financial. Children are worth so much money to the school. So, the more children the more money. Success breeds success. But what about schools that are in failure? So, my view is that it is misguided. It leads to a separation of quality, not to a general improvement.
>
> Umm, I think, using market forces does not improve the system as a whole. How does it improve weak schools? It is the weak schools that we should be interested in, not the strong schools. Because those are children who are not getting an adequate education. And simply closing weak schools and reopening them as academies doesn't work. So, the market forces itself and pupils being worth so much money is a misconception as to how quality is achieved. The idea is that competition between schools for children drives up quality. But does it? Where's the evidence? (Cooper)

Orion also referred to the damaging impact of a system based on 'market values' because this system 'privileges some [children] but damages many others.' In a similar vein, Dennis questioned the idea of school improvement through parental choice as 'only partially true'. As he put it:

> Well, the government certainly believes that if you give parents choice based on the results of inspections or tests, then that will improve the system. That's what the government's view is. And, in a limited way that's true. If you judge the system simply in terms of outcomes, then publishing the outcomes of those tests or those inspections is likely to make parents more sensitive to schools that the government at least regards as good or otherwise... The government's view is that parents place a great deal of emphasis on test scores or inspection reports. I don't think that's actually true. I think, the majority of parents are more interested in whether a child is settled, happy, safe in the school rather than necessarily that their score very highly up on a test or whatever. (Dennis)

Overall, therefore, the participants' views on parental choice converged around the theme of competition as a 'false God'. As Kelvin put it, parental choice 'has been this great God that we all bowed down to'. However, 'it is a false god' (Kelvin), a 'myth' (Felicia), a 'lie' (Nora), 'misleading' (Cooper, Felicia) because 'you can't always guarantee' parental choice (Kelvin).

Ned pointed out the fact that 'parents don't particularly have a choice, but they are able to express a 'preference' for which school they might like to go to. For Orion:

> choice is an interesting issue. And if we could genuinely choose what kind of education our children got, we could genuinely make an informed decision… Some of us would send our children to schools that specialise in producing football stars and some of us could send our children to schools that produce mathematics stars. If it really worked like that; then, maybe choice would be a good thing. But it doesn't work like that. (Orion)

Kelvin, Nora, and Orion also paid attention to issues of parental choice such as the 'urban versus rural and suburban divide' (Orion). There might be a sufficient number of schools in London to choose from but 'not everywhere in England is like London' (Kelvin). In this regard, they critiqued 'London-centric' (Orion, Kelvin) policies and argued how, in rural areas, 'the geography makes the choice' (Kelvin) and either 'you go to your local school, or you spend a lot of time on a bus' (Orion). Thus, there is not much choice in the system. As Kelvin put it:

> A big problem with lots of education policy in this country, is it's all made up in London. And you'll realise from where we're having this interview that not everywhere in England is like London. So, the first school I led was half an hour in any direction from any other schools. This idea of choosing one school over another because of its Ofsted report was meaningless because geography dictated that you went to your local school. However, in the second school that I went to, there was another school half a mile away and another school two miles in the other direction. So, it was much more along those lines. (Kelvin)

Further factors that enable or constrain parental choice included: resources (time, money, connection); awareness of choice and knowing how to make a choice. For instance, Nora argued that whether parents might be able to 'position' their children 'to go to a good school' depends on 'the money and… connections'. For Nora, parental choice encourages playing the system when they 'move to a new house' or 'have to lie to pretend that they live somewhere in order to get their child into a particular school'. Further, she said, 'if you want to, you can apply to go to one of the other towns. But it would mean every morning, you'd have to get on a bus at your own cost and travel maybe 25-30 miles to go to school' and then she questioned 'So, why would you do that?' Similarly, for Felicia, choice depends on factors such as 'enough understanding of the education system to know that there is a choice in the system' and being 'able to' send their child to a school out of the catchment area. As she explained:

> I think, parental choice is a myth. I don't think, there is any such thing as parental choice. I think it depends if you have enough understanding of the education system to know you even have a choice. I think, it depends on what you're basing that choice on. I think that it advantages parents who are able to get their child to a school out of the catchment area. How do you even get them there if you work? It's not even a marker of poverty. I'm a working mum. I work full time. I don't have parental choice because how do I get my children to that school and back again? I can't. So, actually, I don't have a choice. I

132

might theoretically. So, I think, it's very misleading and actually untrue. I don't believe, parental choice exists. (Felicia)

Since the practicalities of choice depended on a range of factors, parental choice 'enhances inequality' (Torr) and 'advantages parents who are able to get their child to a school out of the catchment area' (Felicia).

The tools for generating information about schools for parents such as league tables and Ofsted grades also received serious critique as both unfit for purpose and leading to negative consequences. For instance, Cooper argued that league tables typically serve 'middle class parents who are interested in their children who will pick a school that gets better results'. He questioned the regime of league tables as it 'doesn't produce creativity… [and] does not produce quality' but creates 'fear'. Ned noted how 'people act not in the interests of the child but in the interests of the organisation so that the organisation can move up to the league tables.' In a similar vein, Dennis argued that league tables are 'distorting' and fail to promote 'teacher professionalism or the improvement of education'. Orion expressed his 'concern' about league tables due to their unhelpful 'overall impact':

> What concerns me is that the overall impact [of league tables] seems to me to not help the course of education and not help the young people receive a better education than they would have if they weren't there. If you have school choice, then you have to have some way of providing parents with information to help them choose between schools. I would be inclined to the view that school choice is given more priority than it should be. (Orion)

The participants also challenged the official claims that grades are to inform parents and problematised the use of Ofsted grades which 'destroy' headteacher's careers and lead to school closures. Felicia noted how Ofsted grades can help parents but also plant a seed of doubt about schools where they send their children. As she put it:

> if we're saying that school evaluation is for parents; well, my children's primary school has been Ofsted inspected and has been put into requires improvement. Now, I'm from within education and I understand what that means and even I'm thinking: 'Well, do I want them to stay there?' And then, I hear the teacher voice in me going: 'No, no, that's ridiculous. It's fine. The school is fine. Ofsted have just found something they don't particularly like. My children are happy. My children are learning. They love that school. Stop being ridiculous.' But it was... I found it quite interesting that I heard that voice in my head go: 'Oh, does that mean the school is bad?' No, I don't think it is bad. But immediately it sets that seed of doubt. And who's that helping? (Felicia)

'Attaching one adjective' to a school was found to be 'absurd and anti-scientific' by Dennis, whilst Nora said that this is 'trying to simplify a system that isn't simple'.

The second set of problems with marketisation strategies was that England ended up with 'is a disintegrated system of individual schools behaving like private schools, with no system locally' (Cooper). Increasing marketisation of schools in England was observed by the participants particularly after 2010, with academies and specialist schools. As Felicia put it:

> since 2010 I think the specialisation of schools is increasing. There seems to be more and more schools that don't fit this kind of community comprehensive model where they offer a narrow curriculum, highly specialised in some aspect of education and the free schools as well… I don't think that's a positive thing. (Felicia)

Dennis noted how 'market-based strategies make it very difficult to plan an overall system for education' and 'some of the chaos or the malfunctions we've got in our English system at the moment are because of an undue emphasis on market-driven strategies'. He stressed the importance of 'national planning' and argued that 'market strategies are a recipe for individualism, lack of coherence and, I think, poor governance of education'. In a similar vein, Orion refrained from saying 'education system' and referred instead to 'a patchwork of education provision' fragmented because of marketisation:

> We used to have an education system and now we have a patchwork of education provision. And that's because of marketisation. So, we have individual providers, you go ahead into the market but apparently choose provider. (Orion)

Nora and Felicia problematised taking schools out of local authority control and creating a range of schools, ostensibly to diversify parental choice (i.e., specialist schools, academies, free schools). According to Nora, in the last 10 years schools 'detached' from their local communities, particularly academies and free schools. Nora's views were echoed by Felicia's experience of working in a specialist school (a University Technical College) which 'has existed purely because of those marketisation policies'. According to Felicia, in such a school 'nobody had any kind of pride or ownership of the school because it didn't really belong to anybody'. Therefore, parental choice and marketisation may destroy the sense of community:

> A specialist school that has existed purely because of those marketisation policies. There is no sense of community. There's no catchment area. You can come from anywhere. And it was quite a soulless place to work. There was nobody had any kind of pride or ownership of the school because it didn't really belong to anybody. And that's why, I came back into more traditional comprehensives like this. Because we are firmly rooted in our community. And I think the danger of parental choice and marketisation is that you lose that sense of community in a school and then I think you're a bit sunk. I think that's dangerous. (Felicia)

In a similar vein, Kelvin argued that the 'school is the hub of its community' and they are not only for 'providing a particular kind of education for pupils'. Schools are also about the provision of adult learning or sporting facilities, drama opportunities or community events.

Kelvin was critical of the 'survival of the fittest' understanding of the market in education, because if the school is 'not one of the fittest it can go'. He noted that 'if as a result of… market-based approach, that school disappears… you are looking at the loss of a hub of the community'.

Another negative consequence of schools opting out from local authority control was the loss of local 'support' (Ned). Ned, who works as a lead headteacher in his local authority, referred to a number of examples of how a Multi Academy Trust causes 'tension', especially if the Multi Academy Trust (MAT) is spread over a large region and, therefore, 'the support team is spread across the country'. To Ned, this is 'complete nonsense' that impedes improvement across the MAT and, in addition, makes it difficult for MATs to cooperate with other local schools. As he put it:

> I work a lot with the local authority in [details removed for ethical reasons]. We have real issues with children who are excluded from school. Because they've been excluded, there is nowhere else for them to go. Because they may come from a Multi Academy Trust that doesn't cooperate or work closely with the other local partner schools… Where you have for example a Multi Academy Trust and a school let's say based in Newcastle and another school based in Birmingham and another school based in London, the support team is spread across the country. And to me, that's complete nonsense and it does not help in enabling school improvement. (Ned)

According to Cooper and Torr, marketisation-based policies distilled the budget and power of local authorities since 1988, and even more significantly, after 2010 reduced local authority control over education. Cooper pointed to an increasing centralisation through academisation, but without the accountability of MATs, and argued that this is a 'self-defeating change' that has been 'destroying any capacity locally to do any work with schools.' As he put it:

> If all schools become academies, what's the difference? You're back to where you started… Making them academy doesn't change anything. All it has done is to take money away from local authorities who now have no ability to do anything at all in schools. Because when a school becomes an academy, it takes some of the local authority funding with it. That's the way the money works. So, what that means is that you're simply destroying any capacity locally to do any work with schools. Local authorities have no money, literally no money. (Cooper)

Torr argued that the 'political motive' underlying converting schools to academies is to control education. As he explained:

> there's a political motive for that. Because local education authorities have typically not been right wing. They've not been conservative supporting local authorities… So, conservative governments have had a tendency to do things that reduce the capacity of local authorities to control education system. (Torr)

Cooper and Nora similarly problematised academies as they are 'businesses and are not even

under educational law apart from school admissions' (Cooper). Cooper described academies as a 'travesty of the use of public money'. Nora problematised 'allowing those business-people to be educators' and expressed her concern that in many academies and free schools 'you're getting principals, so directors of schools, who aren't teachers.' She highlighted the importance of teacher professionalism and argued that being 'really good managers businesspeople etc. doesn't make them a good educator' because they do not have 'underlying knowledge about what makes a good lesson or what makes a good assessment or whatever it might be'.

Overall, these experts problematised parental choice because it is not only geographically restricted but also restricted to the resources parents have such as time, money, network connections, as well as parental awareness and knowledge of how to make a choice. Parental choice is a tool which can never be implemented truly, and a 'vote winning' exercise of politicians. Also, applying the market idea of consumer choice with the competition philosophy in education is a 'misconception' (Cooper) in public education and more fit within the business context. This conception does not improve schools but creates 'aggressive war' (Kelvin) between schools, with winners and losers (Felicia, Kelvin). It acts as a 'recipe' (Dennis) for 'individualism' by relying on individual school improvement, leads to a lack of coherence and fragmentation in the system, as well as 'poor' governance of education (Dennis). Market-based strategies are a key reason why it is very difficult to plan an overall system for education in England. Therefore, competition and applying market principles in education is a severe problem. Parental and wider community involvement suggested by these experts is discussed in Chapter 7.

## 6.4 Conclusion

The interview data discussed in this chapter revealed a strong polarisation in the views of experts, whereby the same *pillar* of the official approach was assessed either as problematic or a strength in the evaluation of schools. The benefits and problems cited by the experts were broadly aligned to their predominantly positive or predominantly critical views on the official approach to school evaluation promoted in England since 1988. The official approach was seen as beneficial by some experts because it was fit for the purpose of driving improvement at the school and national levels. However, some experts saw it as problematic because of its key focus on *individual* school improvement, instead of system-level improvement. Also, whilst some participants saw national standards and frameworks as beneficial in making people act in a 'right' direction, some participants problematised how and why standards were used.

Specifically, implementing standards rigidly, designing them prescriptively and using them to judge schools and as tool of social and political control were seen as problematic for creativity, innovation and teacher professionalism. Therefore, whilst all educational experts who participated in this study referred to having national standards as essential, their views on 'how' and 'why' to use national standards differed.

A similar polarity of views was articulated in relation to accountability. The benefits cited by some experts included Ofsted inspections and high stakes accountability as tools for 'bringing people into line' or pushing them to improve. In contrast, the problems with accountability cited by some experts were mainly with how it was conceptualised (as external accountability 'from above') and enacted (to sanction and punish schools which did not deliver the expected standards). In relation to the punitive nature of its enactment, the problem was not allowing schools to answer and justify their 'unexpected' or less successful outcomes. In relation to conceptual problems with accountability, equating accountability with Ofsted inspection was cited as a misconception. The idea of improving schools through accountability was also deemed problematic because it is based on an assumption that schools need an external driver to improve. Another conceptual problem with accountability was its outcomes-based orientation which concentrates mainly on quantitative student performance data, narrowing down broader educational purposes and creating 'perverse' incentives for schools.

In relation to parental choice, the benefits of Ofsted reports and school league tables were cited by some experts as being transparent to parents and providing them with the 'right' information. Because parental choice is their legal 'right', they should be able to make informed choices. However, some experts problematised transparency framed in terms of parental choice by pointing out that it is restricted to the resources that parents have such as time, money, connections, as well as parents' awareness and knowledge of how to make a choice. Parental choice was, therefore, interpreted by some participants as a tool to win 'votes' from middle-class parents, which can never be truly implemented. Another problem cited by some participants was marketisation strategies and the competition philosophy behind the official approach to school evaluation. Designing evaluation around competition in public education was flawed and created an 'aggressive war' between schools, with winners and losers. Whereas all experts valued parental involvement, they differed as to whether parents were considered as 'outsiders' or 'insiders' in the education system. This finding will be explored further in Chapter 8.

# CHAPTER 7: ALTERNATIVES TO THE CURRENT OFFICIAL APPROACH TO SCHOOL EVALUATION

## 7.1 Introduction

This is the second of two chapters of this thesis focused on the analysis and discussion of the empirical data collected through semi-structured interviews with educational experts. The chapter focuses on research question 2 by discussing alternative approaches to school evaluation, which the participants considered as a more appropriate foundation of school improvement. My analysis of interview data pertaining to alternative approaches focused in particular on how the participants appeared to position themselves in relation to the four *pillars*, constructed in the official approach as key 'drivers' of school improvement. Two key positions emerged from the data: 'pragmatic' and 'radical', discussed in Section 7.2 below. Although these two positions have been identified from the patterns that emerged from the interview data, due to the number of participants in this study (see Chapter 4, Section 4.7), I am not presenting these two positions as generalisable to a broader population of educational experts or education professionals. Instead, the distinction between the pragmatic and radical positions which emerged from the data seeks a more complex set of views than the distinction between advocates and critics of the official approach to school evaluation discussed in Chapter 2. The official approach has been based on hierarchical, 'top-down' relationships which position school leaders and teachers as 'policy takers' rather than 'policy makers' (Gunter and Forrester, 2009). However, what also matters is how educational experts and education professionals working in a range of roles assess approaches to evaluating the system. In order to improve schools, we may need both pragmatic and radical assessments, depending on the context. It might also be useful to engage groups of experts who hold different positions in dialogue.

Section 7.2 explains how the two positions (pragmatic and radical) were derived in the process of data analysis. Section 7.3 presents suggestions for improving the current official approach by participants who appeared to take a pragmatic stance when asked about alternative approaches to school evaluation. This is followed by a discussion of alternatives offered by experts who took a radical position in Section 7.4. Section 7.5 summarises the key findings on alternative approaches to school evaluation from both positions.

## 7.2 Identifying the pragmatic and the radical positions

This section explains how the two positions emerged from the interview data collected in

response to the question about alternatives to the current official approach to school evaluation (Appendix 2). The terms 'pragmatic' and 'radical' capture two positions which were derived in the process of data analysis and cross-checked by referring to how these two positions are defined and conceptualised in published sources. With regard to the pragmatic position, the Oxford English Dictionary (OED, nd) defines pragmatists as individuals who are guided by 'common sense', who focus on the current circumstances and are likely to suggest moderate adjustments within a balanced, 'sensible' view rather than radical critique. The adjustments might take the form of 'incremental' improvement, as defined by Lindblom's (1979) incremental theory:

> incremental decisions involve limited changes or additions to existing policies... This approach recognises the less than ideal circumstances under which administrators must make policies. There are very real limits of time, brains, money etc. on administrator's ability to understand complex problems and make different policies about them… Incrementalism is also realistic because it recognises that policy makers lack the time, intelligence, and other resources needed to engage in comprehensive analysis of all alternative solutions to existing problems. Moreover, people are essentially pragmatic, seeking not always a single best way to deal with a problem but, more modestly, something that would work. (Lindblom as cited in Anyebe, 2018: 14-15)

According to Lindblom (1979), incrementalism does not always mean 'small' increments, on the contrary, it may refer to 'large' increments. Of essence to incrementalism is that the change is not radical. Incremental change can be seen as a continuum, as for example in the case of suggested alternatives that are 'in line' with the current system rather than radically departing from the established direction of travel.

A radical position, on the other hand, is more likely to be guided by a vision which goes beyond the current system and takes a radical stance, for example by questioning the very *pillars* of the system (Chapter 4, Section 4.5). Alternatives proposed from a radical stance are, therefore, likely to include forms of evaluation which are methodologically, ethically, and epistemologically sound and justifiable, but which may be unrealistic in terms of resources. This position was captured by Nora's reference to a 'magic wand': 'if I could hold a magic wand, what I would do is just make it all equal… everyone gets to go to a good school'.

Looking at each interview transcript, I have sought to ascertain the position that each participant appeared to take in answer to interview questions about alternatives to the current official approach to school evaluation. The participants who appeared to take a radical stance, questioned the very foundations of the current approach. They referred to the system not being 'right' (Ned) and in need of 'resetting our targets' (Cooper) and 'removing' the risks and threat

of accountability as one of the 'biggest barriers' (Felicia), as well as improving the 'experience of the child' (Ned). These participants advocated collaborative, local and community approaches to improving schools and a holistic development of students within an overarching ethos of trust in professionals. Table 7.1 illustrates the radical stance taken by the participants in relation to alternatives to the current official approach to school evaluation in England.

| Participant | Statement illustrating a radical position |
|---|---|
| Cooper<br>Local authority manager | We ought to reset our targets around how we should be educating children for the modern world, not just a pretend academic system that the middle classes all aspire to. |
| Dennis<br>HMI inspector | the findings of any inspection are subjective, contestable, and never definitive… a 'snapshot'… the time specific… |
| Nora<br>Researcher | [Parental choice] has been the biggest sort of lie that is sold to the people and still continues to be. So, if I could hold a magic wand, what I would do is just make it all 'equal'. Say, everyone gets to go to a good school. |
| Gordon<br>Researcher | the local and the national inspectors need to work hand in glove with each other. But we have got rid of them because of austerity… They could be brought back in. |
| Orion<br>Researcher | I'm rather in favour of professional control. So, you have a National Board of Education and it's staffed by professionals… educational administrators, senior teachers and subject specialists, educational psychologists, and child welfare specialists and so on and so forth. Actually, staff it with professional people. And they have an agreed budget for five years and they put papers to Parliament for the funding of education. But they are not themselves politically controlled. |
| Kelvin<br>Headteacher | I don't think government should set education policy. Government don't set monetary policy… Why not education?... de-politicise education and have a group like the monetary policy committee nationally which reviews education… |
| Torr<br>Teacher | I'm unconvinced by that hierarchical process… accountability should work entirely the other way… I genuinely believe that as a teacher I should be accountable to my students and not to my managers. So, school leaders should be accountable to teachers. Everyone should be accountable to the user. |
| Ned<br>Headteacher | The system isn't right. Because at the moment, it is still too top down, it is still too linked to government kind of ideal world view like this is a good school rather than linked to the experience of the child. |
| Felicia<br>Deputy head | Evaluation can never work because the risks of accountability are so strong. I think that's one of the biggest barriers. We have to remove that… threat of closure or loss of jobs hanging over their head… I think, not trusting teachers -us - is a barrier |

**Table 7.1:** Statements illustrating a radical position

By contrast, the participants who appeared to be more orientated towards the pragmatic position referred to some additional improvements and 'tweaks' (Eduardo) to the current approach to evaluating schools. Accountability and standardised national inspection are for these participants the two fundamental *pillars* which cannot be replaced or abolished. Similarly, published performance tables are vital in informing parental school choices and driving school improvement. For instance, when asked about his 'suggestions for improving the evaluation of schools' (Appendix 2), Kent highlighted an 'evolution' of the inspection

system and suggested two further minor improvements. Firstly, cancelling the 'proportionate inspection' policy for outstanding schools and, secondly, inspecting all schools more frequently:

> Honestly, I'm not an advocate for Ofsted. I am not in their employ. But I think they've got it right. You know, I think, because it's changed. We haven't had a static model for 30 years. It has changed. And I think that evolution has been right. I think, it has been done in the right way. And I think you know, the learning points for Ofsted are about the quality of teams and their inspectors as opposed to the framework itself. So, there is no… anything other than the two things that I've mentioned, I wouldn't change it. (Kent)

Importantly, at times these participants did refer to more far-reaching suggestions. However, these suggestions were immediately followed by claims that they would not work. The barriers to any radical change cited by the participants were to do with the inherent need of some schools to be 'pushed' to improve (Bella), resource limitations of Ofsted (Eduardo, Neil), and political interference in educational and Ofsted policies (Bella). These suggestions appeared to be put forward by participants to `highlight their awareness of counterarguments and critiques of the official approach by some stakeholders. For instance, both Bella (from a pragmatic stance) and Kelvin (from a radical stance) pointed out that 'political interference' is a challenge and barrier in the English education system. However, their suggested alternatives differed. When asked directly, Bella appeared to be reluctant to suggest a way to deal with this challenge. Instead, she said that 'political interference… is a barrier' but 'not that we can do anything about it, because that's the system in which we live'. On the other hand, Kelvin suggested the need to 'de-politicise education and have a group like the monetary policy committee nationally which reviews education…'. As he pointed out, 'government don't set monetary policy… Why not education?' Table 7.2 illustrates a pragmatic stance.

| Participant | Statement illustrating a pragmatic position |
|---|---|
| Richard Policymaker and inspector | …schools must report themselves on how well they are doing. I think, in an advanced system that would be a good model. But I would still argue you would require an external objective evaluation sitting alongside that process. Because 'we all think we're better than we are'. That's the truth. |
| Bella Inspector | The sensible school embraces Ofsted and actually uses it as a tool for their own improvements. |
| Neil Policymaker | …the stakes around progress are too high… need to be lowered. But … it's still reasonable to publish performance tables about the results… Because that is… what you're getting for the public investment in having a school. |
| Eduardo Researcher | Where you have systems where you give schools a lot of autonomy; then, you need alongside that accountability mechanisms… |
| Kent Principal | Why get ants in the pants about it [publishing school performances]? You know, it's just the way it is… You can't have an inspection system based on trust. How can that work? I'm not being funny. But you know, give me an inch, I'll take a mile. |

**Table 7.2:** Statements illustrating a pragmatic position

This chapter now proceeds to discuss suggestions for improving the current official approach from a pragmatic position.

## 7.3 A pragmatic stance on improving the current official approach to school evaluation

From a pragmatic stance, the participants asserted the importance of accountability as a key driver of school improvement (see also Chapter 6, Section 6.2.2). In other ways, on the issue of accountability, there is no alternative. Consequently, the participants also advocated for continued national inspection which provides 'objective', 'national' and 'standardised' system (Bella, Richard, Kent, Neil, Eduardo). Ofsted, as a source of assuring school quality for the public, provides a steer to people to act on. The participants advocated retaining Ofsted grades for schools, as well as the published performance tables as: the 'end product' of a school's cycle of improvement (Neil, Richard), an incentive for schools to improve and a reliable source of parental school choice. A related suggestion was to train schools in self-evaluation (Eduardo, Bella), based on a perception that some schools do not know how to use data they collected. The key improvements to the system suggested from a pragmatic stance were thus mainly focused on changes to the evaluation methods designed by the Department of Education (DfE) and Ofsted. Participants' suggestions included the following improvements, discussed in detail below: an improved cycle of inspection for 'all' schools (Section 7.3.1); increasing resources for more comprehensive Ofsted inspection visits (Section 7.3.2); more detailed data on views of parents and students to add to the current Ofsted data (Section 7.3.3). An improved cycle of inspection for 'all' schools was predicated on abolishing the 'proportionate inspection' policy introduced by the DfE in 2015 and inspecting Multi Academy Trusts (MATs) as legal entities. The participants also suggested more effective support for professional development in schools (Section 7.3.4). Importantly however, the support for schools was referred to as the remit of DfE because the main function of Ofsted is to 'diagnose' and 'inform' (Neil). This does not mean that Ofsted is not to support schools, but its role is more about setting expectations through the Ofsted framework and inspectors' feedback to steer the system. Resources (money and time) were seen as the main reasons why Ofsted cannot visit schools more frequently. Although participants from a pragmatic stance noted 'political interference' and resources as two key barriers to improvements in policymaking and Ofsted inspection (Bella, Eduardo, Neil), no suggestions were cited for addressing these issues.

From a pragmatic stance, participants argued that 'all' schools should be subject to scheduled inspections and were critical of the 'proportionate inspection' policy (DfE, 2015). This policy made 'outstanding' schools exempted from routine inspection until 'results go down' (Kent) or Ofsted receive a 'serious complaint from parents' (Richard). Neil, a policy maker at Ofsted, noted that this policy is 'never something that we sought', because some schools have not been inspected for ten years. He also pointed out that this policy caused 'perverse incentives' such as encouraging schools to teach to the test to create outstanding schools based on performance tables rather than a more rounded education. Similarly, Eduardo, a researcher at Ofsted, noted the limitation introduced by the proportionate inspection rule (DfE, 2015), because 'schools don't stay the same'. He stated that allowing Ofsted to inspect ten percent of outstanding schools is 'not very much':

> So, there are things like the fact that we cannot easily inspect schools that have an outstanding grade that can be a problem. Because then, you see that some schools are not being inspected for ten years or something like that and of course as you know schools change in ten years. We probably have got quite a number of new people in the school for example leadership and teaching levels. So, is that school really still as good as it was 10 years ago? We cannot be certain of that… (Eduardo)

In a similar vein, Kent and Richard suggested abolishing the exemption from inspection policy for outstanding schools for two reasons. First, even with the results not going down, staff well-being, the curriculum, and teacher-student relationships 'could be appalling' (Kent). Second, some outstanding schools that have not been inspected for over ten years may be 'sitting on' (Kent, Richard) their outstanding judgements. As Kent explained:

> My biggest bugbear is that non-inspection of outstanding schools. That really aggravates me beyond belief. Because I feel that they should be subject to scheduled inspection as everybody else is. You know, my own children school hasn't been inspected for twelve years. How is that fair? And I can tell you now, that school is not outstanding from my professional view… You know, if Ofsted say, actually we're going to ensure that all schools irrespective are going to be subject to Section inspection every five years, that will be fine. That for me would make it a level playing field. At the moment, it's not…
>
> There was a recent case of an outstanding school. This outstanding school turned out inadequate because of safeguarding. And that school has been outstanding for a number of years. So, you're telling me, safeguarding is suddenly deteriorated? So, those children were in a school where safeguarding was inadequate and a number of children over a number of years have been in a school where the safeguarding is not adequate. (Kent)

Kent's and Richard's suggestions were echoed by Bella, though she appeared to be more moderate in her assessment of the DfE exemption policy. Specifically, she found it 'very sensible' to have a 'mechanism' for inspecting schools more or less often. The timing of Bella's

interview may be of importance here, as the interview took place before Chief Inspector Amanda Spielman's speech announcing Ofsted's wish to inspect all schools (Roberts and Hill, 2019). Bella suggested an inspection schedule which covers 'the lifetime of a child in that school':

> I think, it's very sensible that there is a mechanism for inspecting schools that need inspecting more often and inspecting schools that don't need inspection less often. You know, if a school is judged to be outstanding and all the indications are that… it continues to be outstanding; then, there's no reason really why Ofsted need to go in and tell them again that: 'it's outstanding'. Even if it's only a short inspection, I believe that there should be an inspection of a school within the lifetime of a child in that school… I worry that we are letting some schools coast a bit because they've been outstanding a long time ago. Particularly small primary schools. Because small primary schools change very rapidly. You know, staff change and therefore the quality changes. (Bella)

Eduardo also questioned the practice of not inspecting MATs as legal entities, similar to inspecting local authorities, especially that a MAT has 'more power in its organisation than a local authority':

> At the moment we are not able to inspect the actual Multi Academy Trusts and that's a little bit strange because we do inspect local authorities. And in fact, a Multi Academy Trust has more power in its organisation than a local authority. Because it's under a different legal basis. Because actually legally the Trusts is the organisation that is the legal body and not the schools. The Trust can delegate powers to local governing bodies in school, but it is the Trust that is 'do you need of accountability and do you need of funding etc.' So, ideally, it would be good if we could inspect MATs. (Eduardo)

### 7.3.2 Increasing resources for more comprehensive Ofsted inspection visits

The insufficient frequency of Ofsted inspection visits was framed from the pragmatic stance as a 'resource' issue (Richard, Eduardo, Neil, Kent), both financial and time. Neil, who works for Ofsted, reflected on the length of inspection visits getting 'smaller and smaller'. He noted that, when Ofsted was first established, inspectors relied on 'qualitative measures' and would 'crawl over every subject, every aspect of the school' with 'up to 44 inspector-days'. Although he found this length of inspection 'extraordinary' and not 'sustainable', the present approach was also problematic, for example, a primary school today may have 'one inspector for one day'. Forty-four days have thus been brought 'down to one', whilst it 'should be somewhere in the middle'. The present system prevents Ofsted from completing a 'thorough diagnosis' because they 'don't have the time to get under the skin in most subjects.' This has led to a focus on the curriculum rather than a more rounded evaluation of the school:

> I'd pick up the story in 92 because that's when Ofsted came into existence. And, at that point, the conception of accountability was to have two balancing arms. 'Performance tables' which are the accountability to parents about 'is the education in their school good enough and are their children getting good results?' And Ofsted which was qualitative, which was 'what is it like in their school?'. And it seems extraordinary but

an inspection in those days would use up to 44 inspector-days and we'd crawl over every subject, every aspect of the school and it would be qualitative. And the two were meant to keep each other in balance.

(…) Should inspections last 44 inspector days? 'No.' Should they last one inspector day? 'No'. Should be somewhere in the middle. And it's absolutely right to slim down our inspection. We've slimed it down too far. And a big part of that is financial pressure. The last 10 years have been austerity and we have suffered, our real terms budget has halved over that period. So, that is an originally positive change that became negative.

(…) So, right now, we don't have the time to get under the skin in most subjects and school and inspection. And we think the curriculum is the most important thing. The curriculum is expressed through subjects, and we don't have the time to get into a lot of them. (Neil)

Eduardo and Richard pointed to the austerity-based reduction of Ofsted's resources and the need for compromises. As Eduardo explained:

Obviously, we have been for quite a long time now in a period of austerity. Unlike other parts of the education system, we have been affected by that. Our budget has been cut by about 50 percent. That obviously has consequences in terms of things like how often and for how long can we inspect... So, there are compromises that have to be made. (Eduardo)

Kent and Richard pointed out that length and frequency of inspections is directly linked to the funding available to Ofsted. Kent suggested a pragmatic solution of two days' inspection followed by a one-day inspection as a 'sense check' to ensure that quality has not 'gone down':

First of get rid of exemption. You know, inspection is every five years, and you know the inspection regime is only led by the resources of Ofsted. It's not there for any other reason. Just we don't have enough inspectors to do more. All the money to appoint more inspectors to do more. But I think, more frequent inspections, you know maybe even if it is just one day. You know, you have two days' inspection and then midway through you get a one-day inspection which is kind of a sense check that things haven't gone down the toilet… Yeah, get rid of the exemptions because that's ludicrous and wrong and decrease the distance between inspections. (Kent)

### 7.3.3 More detailed data on the views of parents and students to add to the current Ofsted data

When probed about the role of pupils' and parents' views in Ofsted's evaluation, the participants admitted that school inspectors could increase parental views 'as an information source for the inspectorate' (Eduardo, Bella, Richard). However, parental involvement was also considered to be time-consuming for parents and unfeasible for Ofsted, due to the current length of inspection. Bella for instance pointed out that 'people seem to forget that parents and students are the real customer' and their involvement is 'essential'. She noted that it is pupils 'who really know the school… and tell you lots and lots of things about the school'. However,

gaining their views could be time-consuming:

> People seem to forget that. They're [students] the real customers... And, I think, it's essential that both pupils and parents are included in the evaluation process. I mean, Ofsted have a process whereby parents can give their views and their names or anonymously. And that, that is really important. And, as an inspector, I would look at those parental views and if there were some key issues that were coming out of it; then, I would be investigating those key issues with the school. Obviously, you're a professional. So, you understand that there's going to be one or two people who just hate everything about everything. And therefore, you know, they, they might write really awful evaluations. But if you're a professional, you can see through some of those or you can say to the school 'tell me about this particular person' and they'll say: 'Yes, there has been an issue'. But, I think, it's absolutely essential.
>
> And, a number of times, I have run parents' meetings within Ofsted and also in other roles that I had where I brought parents in to discuss a particular issue during the evaluation of the school. And it's crucial, it's really, really important. Because yes, they are the customers, although of course, the pupils are the customers. And I think, we forget… So, yes, I think, parents need to be involved as much as possible, as much as they want to be. (Bella)

Neil argued that involving parents properly is 'a relative weakness in our design' and 'should have a much bigger place for parents'. However, due to the length of inspection, 'it's really hard for us to do successfully'. On the other hand, a greater parental involvement was approached with caution, first because of paucity of 'evidence' on the role of parental involvement in schools' decision-making and improvement. Second, greater parental involvement may be in conflict with the broader societal values that education needs to instil in children. This finding is discussed further in Chapter 8, Section 8.2.3.

Overall, there was consensus about the importance of the views of parents and students as a data source for the inspectorate, though resources limited the additional data that could be taken into account.

### 7.3.4 More effective support for professional development in schools

From a pragmatic stance, participants suggested more support for professional development in schools. The support cited as particularly useful in enhancing schools' competence would shift focus from the use of data to making available the right 'treatment' for schools as a response to Ofsted 'diagnosis' (Neil). Such support was deemed to be the responsibility of the DfE or other stakeholders (NLs, LLs and the SLs) in the system, who need to 'show the same understanding about the quality of education' with Ofsted but currently 'they don't'. In his role as a policy maker who works for Ofsted, Neil observed that 'there is a disconnect between our diagnosis and the treatment that's available in the system'. Whilst Ofsted is part of the 'chain'

of support, their main role is more about setting expectations through the Ofsted framework and feedback:

> there is a disconnect between our diagnosis and the treatment that's available in the system… It's, it's, it's been there for a while, and it's not fit for purpose. And so, I'm not sure that the treatment is there currently for the problems that we diagnose. So, those are the things that stop that work. But in an ideal world, that's the functionally... dead… it's the excessive focus on data and it's the insufficient support and treatment for schools that we diagnose. (Neil)

Neil found the state of continuing professional development in England to be 'extremely poor at the moment' because:

> you can get part-time courses, or you can get courses about teaching and learning. It's very hard to get well sequenced courses about pedagogical content knowledge, 'how to teach my subject better'. And that's what we need and that's where the gap is. So, if I had a hundred million pounds, I might put it into CPD... (Neil)

When probed about improving the quality of school self-evaluation, the participants argued that more training and support for school governance needs to be available. For instance, Bella pointed out that self-evaluation is 'an essential part of school improvement' (Bella) but some schools are not aware of its benefits. She also argued that 'not all schools have the time or the resources' to engage in training, particularly in the much-needed training to support school governance and 'how best to self-evaluate'. However, she also emphasised that self-evaluation was not an alternative to 'national inspection' and the 'accuracy' of school self-evaluation needs to be checked by an independent organisation:

> A lot of places, so the big Academy trusts for example. I'm sure that they would like to do their own and they do, do their own evaluation of their schools. And that's good and that's, that's absolutely right. I think, they would like to think that they don't need anybody coming in and checking that that is OK. But I think they do. I think there needs to be a mechanism for checking that school evaluation is accurate. (Bella)

Bella also pointed to what she perceived as the schools' need to use assessment data more effectively:

> There is a massive amount of training and development opportunities about how you actually measure things. So, particularly, the quantitative stuff… the assessments. But there isn't enough in my opinion training and development for schools and for teachers in the use of assessment data. They're very good at measuring and you know as I said that they're all data rich these schools. But they haven't had enough training I think so a massive training implication for how individual teachers and school leaders use the assessment data which they've got. They've got tons of this stuff. But some of them are not using it effectively. So, effectively, in order to improve you like. And the same is true of the quantitative data. It's easy to collect data but it's not as easy to know what to do with that data in order to make the schools better. So, I think that there is a real need in the schools at the moment for how to use assessment data effectively. (Bella)

Another aspect of training cited by Bella related to 'learning exchange' between schools and improving support networks for school leaders. Such networking between schools was exemplified by MATs and private-state school partnerships, 'because of the size of them' (Bella). For example, 'one particular school is particularly good at something like school self-evaluation; then, they can get members of staff on that school to share that with their colleagues in other schools'. This approach was no longer possible within local authorities due to 'cut in terms of finances' (Bella).

Whereas school peer review was deemed 'useful' in supporting school improvement, its use should not replace the national inspection regime nor decrease the need for inspection. This is because this model does not have 'standards' (Bella) and cannot provide 'independent' (Kent) views. Bella referred to a 'place for peer-to-peer review' and its benefits to professional development because of its ethos as an evaluation process 'without feeling threatened'. However, peer review should not replace national inspection because 'you've got to have certain standards that everybody understands and not everybody's understanding is the same'. Bella then provided an example of how outstanding schools in some MATs can help out 'special measures' schools to improve.

Kent, however, appeared to hold less positive views on school partnerships and school-to-school peer review, because such partnerships and networks can turn into 'friendly' conversations, based on a 'personal relationship rather than professional working relationship'. He also questioned the quality assurance of such peer review, which does not address the question of 'who judges the quality of that peer-to-peer support'. As he put it:

> You know, independent view may not make you feel good about it. You don't want the feedback, but you need it. So, you know, you need an independent person to look. I mean, an independent person could be anyone you know whatever, could be Ofsted… [or] a consultant… You know, I don't want somebody to tell me my school is good, if it's not. I just want someone to tell me the truth… peer to peer support, who judges the quality of that peer-to-peer support?

> And I think, very often, it could be that you had the same school improvement partner for so long and this ceases to become objective. You know, it just becomes a conversation. That's not helpful to anybody either the improvement partner or the school that they're going into. Because in some cases, it becomes personal relationship rather than professional working relationship. And it becomes far harder to see the negatives. You know, I give you an example. I like to have a sense check of what's going on around the school and I can say: 'Well things getting better or things getting worse or whatever'. But I can have you know friends who are principals of other schools come in and they say: 'Oh it's wonderful, it's fab.' I don't want that. It doesn't help. So, you know, we'd engaged someone to come in and we know, and we trust. But some of that is not going to pull any punches about what they're saying. And that's important to me. But we

certainly when engaged that person year on year to 10 years eventually they would become complacent. So, it's not going to happen. So, I think it was a waste of time. (Kent)

Some of the themes discussed above recurred in discussions of alternatives offered by experts who took a radical position, with the issues of peer review, accountability, and Ofsted inspections presented in a completely different light.

## 7.4 A radical stance on alternatives to the official approach to school evaluation

As explained in the introduction to this chapter, interview discussions about alternative approaches to school evaluation also yielded more radical answers than those expressed by the experts who suggested minor improvements to the current system. From a radical stance, experts' views seemed to converge on two main problems with the system as it is currently designed and put into practice. Firstly, problems with the evaluation of the system of public education itself. Secondly, and linked to this, accountability, marketisation, league tables and Ofsted inspection set up in the current system as external drivers of school improvement were considered highly problematic (see also Chapter 6, Section 6.3). The current system was described as focused on 'a single school improvement' (Cooper), creating winners and losers due to the reliance on the competition. From a radical stance, the participants advocated for a need to re-conceptualise the *pillars* of the current official approach to school evaluation to serve the 'whole system improvement'. From this standpoint, suggestions included: reconceptualising school evaluation within the 'collective' and 'collaborative' rather than the 'competitive' approach to school improvement (discussed in Section 7.4.1); priority being given to community and 'voice' over 'choice' (Section 7.4.2); reconfiguring 'accountability' within a complex, contextualised understanding of school improvement (Section 7.4.3), and reframing the system of school inspections (Section 7.4.4).

### 7.4.1 Reconceptualising 'school evaluation' within a 'collective' and 'collaborative' rather than 'competitive' approach to school improvement

From a radical stance, participants' views converged on the need to reconceptualise school evaluation within a 'collective' and 'collaborative' rather than 'competitive' approach, in order to improve 'all' schools, at the system level. This collaborative approach would be put into practice through improvement policies focused on the quality of the *system* instead of the improvement of *the single school*. For instance, Cooper, as a managing officer in a 'big' local

authority tasked with introducing the reforms that come along with ERA in 1988, argued that 'the fundamental mistake' of the ERA reforms was 'thinking that schools are separate' and focusing on 'school improvement for a school'. According to Cooper, school improvement 'should be seen as a collective approach across all schools, not individual schools' because 'all schools are connected by the children'. Therefore, he suggested that 'the quality of public education as a whole should be the only thing we should be interested in, not of some schools at the expense of other school, all schools.' In other words, a system that focuses on 'self-improvement through collaboration through changes at the level of the system or operating at the level of the system, not seeing schools as competitors but seeing schools as in the same business with the same children'. The 'fundamental mistake' of ERA 1988 was explained by Cooper as follows:

> So, school evaluation was an important element of those reforms and that was when school budgets were introduced, when pupil-led funding was introduced. And it was portrayed as all being to do with school improvement… That is misconceived... it is fundamental mistake of thinking that schools are separate, but they're not… in a system of public education. And so, it is judgmental and relies on the idea of parental choice driving the quality. And the mistake that it makes is that in a public system we shouldn't be interested in school improvement for a school. We should be interested in school improvement for all schools. That's fundamentally different. So, I think, it's misconceived. I think, it doesn't work, hasn't worked. We've had it now for 30 years and it hasn't delivered what we were told it would and which people like me never believed it would deliver. For the quality of public education as a whole; that's the only thing we should be interested in, not of some schools at the expense of other school, all schools. (Cooper)

In Cooper's view, the current, ill-conceived system based on financial mechanisms leads to a 'separation of quality' rather than a 'general improvement' at the level of the system:

> schools are connected with each other by the children. If you have a system where parents can choose the school they go to, then, by not going to school A and going to school B, school A and B are affected. So, all schools just turn into… in large that to the level of the system… Because our system or mechanism is a fact… is financial. Children are worth so much money to the school. So, the more children the more money. Success breeds success. But what about schools that are in failure? So, my view is that it is misguided. It leads to a separation of quality, not to a general improvement. (Cooper)

For Cooper, operating at the level of the system would mean replacing competition with collaboration:

> And the system that works out on school self-improvement through collaboration through changes at the level of the system or operating at the level of the system, not seeing schools as competitors but seeing schools as in the same business with the same children; those things… the evidence internationally I understand is that those systems tend to work better. For example, New Zealand followed the English model and abandoned it… (Cooper)

Similar to Cooper, Ned, Kelvin, Gordon, and Dennis also asserted that schools should develop all together, learning from each other, within the 'collective' and 'collaborative' rather than 'competitive' approach. 'Peer review' between schools (Ned, Gordon, Dennis), 'peer inspection' (Cooper), 'peer appraisal' (Cooper), 'the critical friendship between schools' (Orion) and 'cross-fertilisation' (Kelvin, Ned) were the preferred forms of school evaluation aimed at 'a broader education experience for students.'

Kelvin and Gordon described their ideal model of evaluation as a combination of school self-evaluation, peer review and area-based inspection. Kelvin, a principal of an academy, highlighted the importance of peer learning through 'cross-fertilisation' as this is an opportunity for school staff to learn from other schools and share their learning within their schools. He noted that this means 'having a throughput staff who come in and encourage people to progress and to move on and then bring in more people in'. Importantly, this collaborative approach would not stem from Multi Academy Trusts but geographically close, area-based partnerships. Similar to Kelvin, Gordon suggested that 'self-evaluation' should be the starting point in school evaluation and then 'local' peer groups of headteachers should look to 'help each other to improve'.

Ned's preference was for an evaluation approach which relies on school self-evaluation and school to school peer review as a 'more creative' and 'more open ended' model of evaluation compared to the 'Ofsted type model', which is 'obviously very paternalistic, very top down and based on a particular view of the way in which schools should function'. Further, he suggested that there was 'no need for Ofsted' if 'really strong local partnerships' are established, and peer reviews are 'rigorous' because a peer review process enables schools to 'really scrutinise each other and support and learn from one another'. As a headteacher of a local authority school, a lead headteacher in his local authority and a National Leader for Education, Ned had highly positive experience of peer review activities between schools in the same local authority (a Co-operative Trust Learning Partnership). He argued that working with 'neighbouring headteachers' and receiving suggestions from them helps to 'move the school forward' and validate school self-evaluation findings. Ned explained:

> I'm interested particularly in teachers learning as well as teachers being evaluated... So, I support the idea that schools need to be evaluated but I would prefer to see a peer review model which means that different headteachers are responsible for working with neighbouring schools to support them to improve rather than seeing an external regulator such as Ofsted operate. My school is part of a [co-operative] Trust of nine schools and within that Trust we all review one another… we work with an external lead who leads the review process. And although that person is Ofsted trained, they

decide, and the headteachers decide on the focus of the review that they wish to pursue. So, it's a very different kind of model to the Ofsted approach.

…I think, what peer review can be is more creative and it can be more open ended. And so long as young people are leaving school with positive attitudes to their future, positive attitudes about feeling empowered and inspired and ambitious for the next stage in their education or work; to me, that's very much the job of schooling. It's about human development, not just academic progression… And unfortunately, the Ofsted model tends to measure purely academic progression, and not human development... So, for example, compared to the rest of the country, outcomes academically in London over the last decade and prior to that with the London challenge have been extraordinarily good. And in parallel, most have been the number of young people who've chosen to go out with knives and have chosen to participate in some of the most violent attacks that we've never seen in our country. So, I would argue we need to look very carefully at a humanising curriculum, and a humanising pedagogy, and a humanising assessment regime that enables young people to feel a sense of self and a sense of purpose for the future. (Ned)

Similar to Ned, thinking about the system rather than a single school in a school 'market' was for Cooper a foundation of an alternative approach to school evaluation, which would be not only more accurate and less 'superficial' but also more challenging to schools:

Start from the system and all schools, you'll take a different view. School self-evaluation is not soft. It's hard. You can hide behind data, anonymised data, might be objective. But is it helpful? So, to get underneath the surface of what needs improving is very difficult. Peer evaluation is difficult. It's intrusive. It's personal. It's far more effective than just data. So, I think, people who say that self-evaluation and self-improvement is soft don't understand. It's difficult… done properly, done well. If it isn't just about having a cosy chat, it is difficult to get it to sit a place where it's effective. So, when we had head-teacher appraisal, I used it to take part in headteacher appraisal. And it depends on a level of trust. How open you are, how willing you are to be honest? That's difficult. But it's professional. So, I think those people who criticise school self-evaluation, self-improvement as soft option don't understand it. They've never done it probably… But in order to get to that place for people to accept that, you have to accept that it isn't about some schools and not all of them… it's about all of the schools and all of the children. (Cooper)

Within Cooper's collaborative approach, schools in a public system are not 'separate' and therefore they:

should work together to improve, not against each other… And that's what I spent my life doing. It is getting schools to try to work together in spite of the system. Because that's far more efficient. If one school does something good, why shouldn't the next-door school know about it? They're not in competition. They're not making cars. They're teaching children. They're the same children. The system, the system is the children. So, it needs to be seen in that way. And I don't agree that parents would think that that was soft or too easy. I think that's a political view which has suited certain political parties to project it in that way. The mistake that was made was that when Labour was in power from 1997, they didn't really change that. They kept that middle-class view. So, what happens is that the middle classes choose and play the system and get their children into better schools. And working-class families tend not to so much. It's just true. So, it works for the middle classes. It doesn't work for those who are relatively disadvantaged, and

those children end up in slightly less good schools. If that was hospitals, we wouldn't allow it. Why do we allow it with schools? (Cooper)

Felicia and Dennis's alternative approach to school evaluation rested on 'a combination' of national school inspection, peer to peer review and school self-evaluation. Self-evaluation should be moderated by peer review and collaboration should be built locally, as also suggested by Cooper, Gordon, Kelvin, and Ned, participants who were also against competition in education. Engaging in more peer-to-peer evaluation can increase schools' confidence in their own judgment. As Dennis put it:

> we now need re-establish school self-evaluation… schools do find self-evaluation difficult. But, if they were to engage in more peer-to-peer evaluations; then, I think, they would perhaps develop more confidence in their own judgments and be able to cope with it. But I don't think we should rely on school self-evaluation alone. I think, it needs to be part of a repertoire of approaches to evaluation as I've just mentioned… in my ideal model of evaluation, there will be inspection along the lines I've just described. On the other hand, there will also be plenty of schools… with other schools to evaluate themselves in terms of what their aims and rationale are. So, there'll be lots of evaluation going on alongside government-imposed inspection … So, we need a combination. But we mustn't overload teachers so much that they have no time to do anything else other than evaluate. (Dennis)

In Orion's ideal evaluation system, what mattered in school evaluation was its overarching purpose, namely 'incremental improvement'. One of the problematic trends that he observed in the current approach to evaluation was attempts to outsource the cost, with self-evaluation used to 'shift the cost from the centre to the periphery', i.e., from the inspectorate to the school. According to Orion, however, the value of self-evaluation is that supports 'incremental improvement', making change sustainable:

> For me, there is a trend in evaluation systems towards trying to outsource the cost. So, reallocate the cost of evaluation from national and funding agency bodies to the providers of the service that's being evaluated. This has nothing to do with the commitment to self-evaluation and everything to do with trying to shift the cost from the centre to the periphery. And if in a sense what's going on with Ofsted is an attempt to reduce the costs of inspection both the financial monetary costs and the costs in kind; then, I'm suspicious. And I'm suspicious because if Ofsted is to do its role properly, it needs to be properly resourced.

> (…) part of the cost of inspection gets transferred from the Inspectorate to the school. For my money, that's a poor use of educational resources… for a primary school or a small secondary school or indeed a large one really, you know, it's a… this disproportionate amount of time gets spent on, I suspect, these self-reporting systems… [are] not a really great tool. It's in part because it's focused too much on the school and not on the operating context in which schools have to work. So, we keep talking about this as if the school is a single and isolated entity, but it works in an environment which others control. And so, the environment beyond the school is an important contributory factor to the performance of the school.

So, it's not that that school self-evaluation is useless. It's just you know suggesting that it plays a major [emphasised] contributory factor to school improvement is, all could do is to fly in the face of the evidence and the evidence is that it doesn't play a 'hugely important' [emphasised] role… So, it is not a magic wand that can be waved, and things can improve. It's just not the case. Okay? But I do think it's the case that you can incrementally improve the quality of education the performance of the school by collecting and using evaluative information, incrementally. And that incremental change over time can add up to significant change. All right?

Now, there are occasions when incremental change is just too slow, and you need a more radical change. But mostly you know, incremental change means they can be absorbed, sustained. It's not of itself disruptive. It's not that suddenly everything has to change. But you say: 'I think that if we could get the parents to spend 10 more minutes a day hearing their children read, it would make an enormous difference. How can we do this? Just 10 minutes.' Or you might say: 'I think that if we were to start school 15 minutes later, you would get more out of the young people half an hour later. Let's try, let's run the experiment'. Because you've noticed that children are sluggish in the mornings. Or you might say: 'We need some way of improving motivation'. And then thinking about it, there is a dozen different ways that you can do this. And then experimenting with these. You know, these are self-evaluative mechanisms. Mechanisms for systematic improvement. None of which would necessarily be the magic bullet but when you add them together. So, I think that where self-evaluation strengths lie is in supporting incremental improvement. So, by definition, it's not going to make an immediate and significant impact but over time it might. The trouble is how do you sustain this. So, you do gradually reap the benefits. (Orion)

### 7.4.2 Community and 'voice' instead of 'choice'

The participants' arguments for abolishing policies of marketisation and parental choice were based on their detrimental effects on system-level improvement that include the fragmentation of the education system and loss of community 'soul' (Felicia). Community support and voice were considered more effective in driving improvement and allowing schools and pupils to 'flourish' (Felicia). Ned suggested that the voices of students give 'a rounded picture of what the school is actually achieving'. Dennis highlighted the importance of voices of students, as well as those of parents, teachers, community, and employers. In place of parental choice and competition over school places, the participants pointed to the benefits of comprehensive, 'common', 'local' schooling under local authorities. For instance, Nora argued that such comprehensive schooling would bring 'equity' to the schools around the country:

It would be a very different picture and I think what you would find is more equity in schools around the country. For instance, learn from Scandinavia. It's just their attitude to what a school is and why it matters and what a good school is, and their notion is that 'all those schools are good'. No one ever says, 'that school has done rubbish and that one was a much better one'. (Nora)

A 'local' school receives support from its community, and, in return, it serves the needs of its community. Kelvin explained how his school was supported by the local community and

became a 'hub' of the community supporting lifelong learning needs of parents and local residents in return. The academy of which Kelvin was a principal 'was in challenging circumstances' and it 'haemorrhaged students on the grounds of [exam] results and… the lowest grade of Ofsted inspection for the longest period ever'. However, the local community 'kind of shrugged their shoulders' to exam results and Ofsted reports and 'we still had more people come in'. This allowed the school to improve over time and become a popular local school. As Kelvin pointed out, because 'we are in the market system', such a school 'could have gone into meltdown and people would have stopped coming' and so that 'schools end up just shut'. It was the support of the community that helped the school to avoid being closed down. In order to involve parents and the local community, Kelvin prioritised 'communication with parents through a range of different formats' and particularly 'free writing'. Furthermore, he also resisted the 'new freedoms' related to academisation such as the option not to include parent and teacher governors in the governing body. His 'fight' to employ teacher and parent governors enabled the school to benefit from 'a very interesting group of governors'. Kelvin pointed out that, due to academisation, 'the governing body is no longer the governing body; the governing body is the trustees of the Trust'. As a result, the local connection may be severed:

> I have got no idea who was on the Trust. Because they met several miles away… didn't come in, didn't see, didn't meet on our premises 'Who are they?'. So, the answer, marketisation is a general concept, one thing, but that is specific aspect of it. I think, you need to have that degree of [local] representation. And some of the measures that have been taken since have actually militated against that involvement of those stakeholders. (Kelvin)

For Gordon, the policy of Michael Gove (Education Secretary between 2010 and 2014) which 'removed parents from the governing bodies' was 'very anti-democratic' because it removes parental voice on issues such as their children's well-being:

> Michael Gove… removed parents from the governing bodies. Parents before that had some role in how the school was run. In that they could vote for two parents to go on to the governing body. Michael Gove has taken them off. Parents are no longer governors of schools. Now that strikes me as very anti-democratic. He didn't want parents to have any influence on what the schools did. He wanted the government to have the maximum influence. So, that is a serious move as far as I'm concerned in terms of the anti-democratic nature of the high stakes system that we've got at the present. Because a lot of parents would be able to say: 'This is creating stress on students.' Again, if you're looking at the British media, you will see there are plenty of reports just now about the high rates of mental problems that young people, particularly young women, are having in secondary schools. And part of that, 'only partly', is the stress of examinations. They have become so important. They've become all important. (Gordon)

Felicia argued schools could 'flourish' more if they were 'firmly rooted in their community'

rather than being left alone and 'belong to nobody'. As she put it:

> What was wrong with the old system? What was wrong with your local school? I don't know why that disappeared. I think that going to your local school... That's how I was brought up. I never entered my head that you would go to a different school. And I went to my local comprehensive and so did all of my friends. And that was our community school.
>
> And having worked in a University Technical College where those schools have been created for marketisation…  A specialist school that has existed purely because of those marketisation policies. There is no sense of community. There's no catchment area. You can come from anywhere. And it was quite a soulless place to work. There was nobody had any kind of pride or ownership of the school because it didn't really belong to anybody. And that's why, I came back into more traditional comprehensives like this. Because we are firmly rooted in our community. And I think the danger of parental choice and marketisation is that you lose that sense of community in a school and then I think you're a bit sunk. I think that's dangerous. (Felicia)

Orion also emphasised the benefits of 'a universal education system where all children went to a state funded primary or secondary school that was their community school' and advocated abolishing marketisation:

> Do we really want children being moved significant distances in order to go to school? My answer to that is unequivocally 'no'. Because that increases the time they spend at school or in transport and, I don't want to do that. In transport seems to me to be dead time. But also, it increases the environmental impact. Why do this? I want children to go to their local school and walk there… Yeah. You go to your local school. Now the consequences of going to your local school is that all schools need to be jolly good. And we need to have ways of ensuring that they are. Once again, I can't see the evidence that marketisation has improved the quality of education. I could be neutral about the policy but show me the evidence. Yes, in one sense, if it works, it works and if it doesn't, we should stop it.
>
> So, my preference, common schooling. Given where we are in Britain; then, choice has a certain inevitability about it. I don't think, that has to be driven by market values. It could and some choice already is driven by cultural values. Now I'm not comfortable with either. But I think, if we accept that a certain amount of cultural choice is inevitable in the UK; I'm not keen on the market stuff. Because, I think, actually, it does… damaged at the system level across for all children. It privileges some but damages many others. It's costly too. It's more expensive form of provision. (Orion)

Whilst Orion pointed out that 'it is very difficult to row back from the school choice', according to Dennis and Ned, going back to a system without parental choice and sending students to their local primary or comprehensive school was 'certainly possible' (Dennis). As Dennis put it:

> it's certainly possible. Because we had it until about 1988. Before that period, then most parents… Well, all those parents opting for the state system as opposed to the independent schools went to their local primary school or their local secondary school. There wasn't much of an element of choice. If they're in the centre of a big city, they might be then able to choose between one or two schools. But mostly, we didn't have a

parental choice system before 1988. And there wasn't very much dissatisfaction by our parents within that. The only element that happened before 1988 is largely was, 'if you are wealthy enough, you could choose to send your child to a public school or an independent school. But, if you were not, you didn't have much choice at all. You usually went to your local primary or comprehensive school (Dennis).

When probed about his views on the diverse system of schooling associated within the official approach with the creation of academies and free schools, Cooper explained as follows:

I do not agree with the premise that we need more 'diversity' through creating different types of school. Where is the evidence that this helps provide every child with a good school? Schools are already diverse because they serve diverse communities. It is not true that schools needed to be freed from local authority 'control', as there was no 'control'. We should trust professionals to educate children and allow them to be diverse in what they do in the classroom. Diversity of governance does not lead to diversity of educational practice. Academies tend to be managerially top-down and allow teachers little room for professional freedom. (Cooper)

The participants' concerns about post 2010 changes to the system related to the formation of Multi Academy Trusts and free schools extended to the potential loss of local partnerships and local ways of working through 'peer review' (Ned) and 'cross fertilisation' (Kelvin). Local support is more practical in terms of geographical proximity (Kelvin), as 'families live locally' (Ned). Kelvin noted that, in some MATS, school staff cannot 'go anywhere near', due to their spread over large geographical areas. In a similar vein, Ned described the policies promoting academies and free schools as 'wrong' and noted MATs cannot easily support the academies under their governance because 'the support team is spread across the country':

I think, schools should be good locally. Because families live locally. And so, to have a Multi Academy Trust that works from one town to another town to another town causes all sorts of tensions. Because if for example, a child needs to change schools, they aren't able to go to another local school whereby they can be picked up and they can be supported. So, I work a lot with the local authority in [details removed for ethical reasons]. We have real issues with children who are excluded from school. Because they've been excluded, there is nowhere else for them to go. Because they may come from a Multi Academy Trust that doesn't cooperate or work closely with the other local partner schools.

So, I believe that we will come to a model which returns us back to geographical based partnerships between schools. Because it's the logical way to operate. I haven't expressed that very well but the issue I think is… in the past we used to have local authority maintain schools. There were many faults with that particular approach. But one thing it did mean was that the support teams were locally based. Where you have for example a Multi Academy Trust and a school let's say based in Newcastle and another school based in Birmingham and another school based in London, the support team is spread across the country. And to me, that's complete nonsense and it does not help in enabling school improvement. (Ned)

Although local schooling was both beneficial and a 'logical way to operate' (Ned), Orion

pointed to the difficulty of involving local authorities into the system again, due to continued 'budget erosion' and the resulting loss of expertise:

> in 1988 local authorities had significant power in education. In 1978, they had 'real' [emphasised] significant power. And there's this decline. And local authorities had both inspection and advisory systems. So, they had advisors in age phases and subjects, and they had inspectors. Sometimes there was an overlap between these two bodies within the local authority. Sometimes they had an inspection service; sometimes they had an advisory and inspection service. And so, local authorities could and did inspect schools. That was in addition to the national inspectorate.

> So, move forward from 78 to 88, local authorities are cash poor and the resources they have for inspection and advice are declining. Get to 1992, they've declined even further. And so, gradually, the role of local authorities in education has significantly declined to a point where… They, now, I think probably lack the advisory and inspection services that would be necessary if inspection would become a devolved responsibility.

> One of the things that the Tory Government did between 1979 and 1992 was the engaging something called 'local financial management of schools'. And local authorities lost money to schools. The financial management of schools became more a school and school governor matter. And this devolution of budgets had an impact on what it was the local authorities could do provide support to schools. Now, this varied across counties as to how they responded to these government initiatives. And some authorities managed to maintain because schools bought back. And in the other local authorities really became impoverished by because the schools went their own sweet way.

> So, the upshot of all this is that although local authorities used to be able to and use to provide advice and support to schools including school inspection; they no longer I think have the resources to do that. And that means that they've lost the expertise. Well, when you lose expertise like this, you can't just find it [a gesture of 'magically'] because you know it's like…This is a bit like saying you know, 'If a hospital lost the capacity to provide accident and emergency services; in order to start that again, they'd have to hire staff, get the right equipment… You can't just switch on like that [a gesture of 'magically']. That's the issue. And finding good staff for inspection and school support is not easy. (Orion)

### 7.4.3 Reconfiguring 'accountability' within a complex, contextualised understanding of school improvement

From a radical stance, participants' views converged on the conclusion that accountability is necessary and there is no 'desirable' (Orion) alternative to accountability. However, to be effective, accountability needs to be reconfigured to encompass a complex, contextualised understanding of school evaluation and improvement. For instance, Orion referred to some alternatives to accountability such as 'an early 1980's Japanese model of 'non-hierarchical view of quality management' that might reduce the 'need' for accountability. However, he also added a 'but', and said that such alternatives might be 'very undesirable'. To him, accountability is needed to 'question' people in any sector including education. Hence, his suggestion was not

to replace accountability with something else, but to create 'more sensitive systems of accountability' that are 'more attuned' to their own complexities. As Orion put it:

> There is no alternative. I mean, I think… Well, actually, there are alternatives but they're not desired. It seems to me that we can't go back to a situation where schools are not really accountable, or accountability doesn't matter. That particular genie is out of the bottle. Let's think in parallel. So, it used to be the case that by large the work of doctors was not questioned. Now that isn't true. No, I can't see us going back to a situation where we don't question doctors. So, with respect to schools, can we see ourselves in a situation where accountability becomes less important? I can't see it myself. Well, I said... I thought there might be an alternative, 'intense standardisation'. So, if you eliminate variability in the provision of a service, it becomes completely standardised. That might reduce the need for accountability. In its place, you might have a form of total quality management. But I mean by that the kind that informed Japanese production lines in the early 1980s. So, it's a kind of non-hierarchical view of quality management. And so, would that reduce the need for accountability? Well, it might. But it would probably be very undesirable. So, I think, what we need more sensitive systems of accountability that are more attuned to the costs and benefits, and the strengths and weaknesses, and the threats and opportunities that accountability mechanisms come with. More sensitive… (Orion)

In a similar vein, Gordon noted that 'in the modern world… everyone must be accountable' and Dennis described accountability as 'a professional obligation the schools should be hassled'. However, all participants also added an important reservation about how accountability is 'designed'. For instance, Gordon highlighted the importance of the forms, accuracy, and reliability of accountability mechanisms:

> Accountability is a good thing. Everyone must be accountable. The two major reasons for accountability in education are: One, it is so important in the modern world. Education is the way that most people will get into good jobs, education, and training. If you come from a poor area education is a main route to a better future. So, that's why, education is so important in all countries. The second reason is we spend something like 30 billion on education in this country. So, the first reason, education is highly important. And second, it's very expensive. And MPs have a right to challenge teachers and say: 'You're getting millions from government. What are you doing with it? So, I have no question about it… But the question then becomes what kind of accountability? What form should it take? How accurate is it? How valid? How reliable? Those become the big issue. (Gordon)

Experts' assessments of how accountability is currently designed, defined, and operationalised in the English education system can be found in Chapter 6, Section 6.3.2. In this chapter, the focus is on four key suggestions for a reconfiguration of accountability presented from a radical stance.

The first suggestion was related to 'what' has been 'valued' since 1988 to render schools accountable. The participants suggested that the accountability system should move from performance and outcomes-based accountability, i.e., student outcomes such as attendance

rates and test results, to a more complex understanding of accountability and broader educational purposes and values such as preparing children for living in society. For example, Dennis said that accountability should 'look beyond just school outcomes' and 'look at the whole life of the school on the qualities and attitudes and skills developed'. This is because 'accountability through test' gives limited information about a school and is not an 'adequate mechanism for getting assessment and evaluation of what a school or what schools can do.' In a similar vein, Orion emphasised the importance of the quality of educational relationships and resources, children's wellbeing, mental health indicators, exclusions, and other factors. Kelvin emphasised the importance of 'wider aspects of education' like resilience and stated: 'I wanted every child who came in our schools to have a responsibility experience, work experience, sporting experience, team experience'. According to Ned, in its current form, accountability relies on 'performance' as an 'easy' way of setting up accountability measures, however, 'ideally', accountability should be looking at 'the long-term success of young people', 'personal development' and 'growth' of students, specifically, happiness, feeling confident, feeling positive about the future, feeling positive about the relationships they can build with others, being articulate and self-assured, feeling that they have a future and a place in our world. Ned also pointed out that:

> there are other ways in which we can help young people to grow and to develop without constantly feeling, 'We've got to test in a formal traditional way by march them into a school or asking them to sit for two hours and recall everything they know about a specific area of human experience'. (Ned)

Cooper, Ned and Felicia proposed a national debate about 'what schools are for' in order to reconfigure accountability.

The second suggestion pivoted on 'whose values' (Ned) and how this should be reflected in standards used to hold schools to account. The participants proposed that accountability should be either contextualised (Orion, Felicia, Dennis, Gordon) or designed locally (Cooper, Ned) against the national standards based on the social, economic, and political context of schools and their communities. Values that come 'from above' and are reflected in Ofsted's standards may neglect the needs of schools, fall short of helping schools and damage professionalism and creativity (see Chapter 6, Section 6.3). The participants who suggested a contextualised understanding of accountability noted that the evaluators should consider the impact of the social, financial, and political context of a school's work. While Gordon argued that contextualised accountability could help 'treat schools in challenging areas equitably', Cooper noted this could encourage the 'creativity' of schools. To Felicia, contextualised accountability

would allow schools to concentrate on their own needs with their own way of improvement strategies instead of the official system that values test results and Ofsted grades. According to Felicia, schools should be able to use their autonomy, regardless of their previous Ofsted grade to 'decide what school improvement looks like for them'. To Orion, contextual variables that impact schools should be made part of a 'story' told about schools in Ofsted reports. Contextual variables are 'important in understanding why the school performs as it does' (Orion), both in terms of success and failure.

Cooper also highlighted the importance of the context within which schools operate, but he was against using national standards to hold schools accountable. Instead of contextualising national standards to school context, he proposed 'locally brokered accountability', designed around local values. In other words, each school should have a unique accountability framework, designed by a local agreement, in alignment with the context of schools and community values. As he put it:

> Schools have to be accountable. It is public money of course there should be accountability. But it should be depending on what you value. So, there should be local agreement about what is valued. So, schools operate in a particular context. So, if I look at the schools around the corner here and I had to think, 'Well what would be its accountability framework?'; then, it might be different from a school 20 miles away. So, it should be locally brokered, it should be by agreement, it should be professional as to what the accountability frameworks should be. But as I say I think having a national inspection system is fine, provided that that's not the accountability system. That isn't accountability.
>
> That should be their accountability and it should be accountability to the local people, who they serve. You can do that. You could do that. And it could be rigorous. You have the national framework of standards you have that in, and you trust the local authorities to work with the schools to produce an accountability system. And if a school is going to go off of the rails too far, you would pick that up through the National Framework if the results were you know… (Cooper)

The third suggestion pertained to accountability 'to whom'. As pointed out by the participants, Ofsted is not the only body to which schools are accountable. For Cooper, to 'secure quality in schools', accountability should be to school governors, that is 'local people', with locally agreed standards for a specific school. However, Cooper also highlighted that accountability to school governors would have resource implications to make it work:

> Local authorities have always tried to train Governors, but it is very hard both to recruit enough people of the necessary calibre to take this on board and to provide enough on-going training. Schools should be accountable to their local communities and Governors are a good way of doing this, but it needs to be taken more seriously and much more money put into making this work. It should not be left to schools.

> You have the national framework of standards you have that in, and you trust the local authorities to work with the schools to produce an accountability system. And if a school is going to go off of the rails too far, you would pick that up through the National Framework if the results were you know…

> Only very few headteachers tell their governors everything. Most governors just rubber stamp what the headteachers tell them. And it's very hard for governors to do anything else. So, I am now a member of the Trust of an Academy. So, I wouldn't have done that for just any academy. I do it for those who I think have got the right idea of a school. But we're trying to make sure that there is real accountability. That it isn't just being told things. Because, you know, if a headteacher says: 'this is true, this is this', the governors of schools have no choice really unless they've been very educated in what goes on in running a school. A few headteachers educate their governors, make them critical, want them to be constructively critical; but very few… (Cooper)

In a similar vein, Ned, and Tor, who were both school practitioners, emphasised their accountability to students and parents. Torr said that his alternative requires a 'radical change' to the current 'hierarchical' structure and suggested that 'accountability should work entirely the other way'. Teachers should be accountable to their students, not to their managers. Ned explained this as follows:

> I think, there needs to be some level of accountability. I don't have a problem with accountability. It's about what form does that accountability take. So, I would say to you: 'I am accountable for the children and the parents who I serve'. So, if a child is unhappy, they come and see me. If a parent is unhappy, they come and see me. And they tell me what the problem is, and together we try and work through to get a solution. But the accountability measures that come above that. They come from the likes of Ofsted, come through the examination system, come from Multi Academy Trusts, come from local authorities. Now I'm quite happy to be held to account by the people that matter, which is the children and if we're not doing the job we should be of course, we need to change, and we need to improve. So, I go back to a peer review process which involves children saying: 'What's working well and what would make school even better'. So, interviews with children about their experience, interviews with parents about their expectations and their experience. So that, you get a rounded picture of what the school is actually achieving. (Ned)

As explained further in Chapter 7, Section 7.4.4, Gordon pointed to the importance of being accountable not only to Ofsted and the government but also to students and community and described an alternative approach would be based on a combination of 'horizontal' and 'vertical' accountability.

The fourth suggestion pertained to reconfiguring accountability as a 'shared endeavour' (Nora) of a wide range of stakeholders, at local and national levels, including both the local community, local authorities and national policymakers. Within this model, accountability went beyond 'being purely school based' (Kelvin), with all stakeholders 'properly' informed about the strengths and weaknesses of policies that impact schools by a 'properly independent

inspectorate' (Orion). In this regard, Gordon noted that the effects of government policies (such as the fragmentation of the system) should be reported to the Secretary of State by Ofsted because the 'duty' of the inspectorate goes beyond simply ensuring that 'schools are implementing government policies.' For Orion, the impact of education policies and funding for schools should also be taken into account, in order to transcend the current, 'unfair' approach whereby only schools 'get blamed' and punished for the impact of the policies.  As he put it:

> The principal things that are missing, firstly, consideration of the political and financial environment in the role that these play in school outcomes… a properly independent inspectorate should hold the Government to account, not just the schools. Or should hold local authorities and the government to account or should hold Multi Academy Trusts and the funding regime or the government to account. But they don't. So, that is the primary missing thing: more attention to contextual stuff. (Orion)

Within this reconfigured approach, accountability should not function as a means of pressure and sanctions such as 'parachuting new headteachers', forcing schools to become part of a Multi Academy Trust or putting them under special measures (Orion). Since 'what the school is able to provide is not unrelated to the wider policy and financial environment', the impact of such policies should be also reported. Accountability is about 'questioning people' (Orion), allowing school staff to 'answer', 'justify' and provide information (Felicia, Torr, Nora). According to Nora, at the heart of accountability is 'responsibility'. She also described accountability as 'a form of assessment' that enables professionals to learn from their 'mistakes', without 'being afraid'. Torr said that accountability should give people 'an opportunity' to explain and justify the reasons 'when something isn't as expected'. Torr's 'dim view' of the current approach to accountability was based on its overarching punitive orientation:

> I have a very dim view of the way that some people view accountability… What they say is: 'You're accountable for this and if it doesn't work, if you don't do what you're supposed to have done, we will sack you'… That's not me being given an opportunity as a classroom teacher to say: 'Well, actually, it didn't work as the way you wanted it to. Because of all of these other reasons.' It just says: 'I've got to do this regardless of whatever else the conditions are'… There's been an increase in the use of the word accountability in the last ten years, maybe the last five years. It's an aggressive hard-nosed approach to managing people. (Torr)

Overall, from a radical stance, the experts suggested that accountability measures should be broadened, from performance outcomes to students' experience and personal development. The school context and local values should also be considered while holding schools to account; each school is different and operates in a different social, political, and financial

context. A contextualised understanding of accountability would help schools to use their autonomy to decide for themselves what should be improved, even if this were to entail rejecting the current definition of 'improvement' based on test results and better Ofsted grades. Not only schools but also wider stakeholders should be held to account, because schools are not the only body responsible for their success or failure. Lastly, accountability should not be enacted as a means of pressure and sanction but rather as an opportunity for justifying the school's practice.

### 7.4.4 Reframing the system of school inspections

Whilst all participants emphasised the need for an 'external eye' (Ned) in the evaluation of schools, this 'external eye' did not necessarily denote Ofsted. If Ofsted were to retain their position, then their role should be 'modified', by acting as a 'professional adviser' (Cooper) or 'professional colleague' (Gordon). The suggestions for alternative approaches to inspection ranged from area-based inspection to replacing routine inspection with inspection to trouble shoot, to inspection based on improved methodology. Importantly, the participants emphasised 'usable' evaluation findings deployed to support the development of schools rather that to use those findings to 'punish' schools through 'naming and shaming' (Bates, 2016). The overarching purpose of external evaluation was seen by Ned as providing 'an umbrella overview' in order to support individual schools in cases when the peer review process exposes some 'fundamental weaknesses':

> if peer review partnerships are established locally and they are proven to be successful; in other words, children leave school feeling positive about themselves, their futures in the world in which they live; then there is no need for Ofsted. If however those peer review partnerships break down or if particular schools need specific support, I believe there does need to be some kind of system which means that somebody is able to step in and say: 'we need to work very closely with this school because the children are being let down'.
>
> (…) Now how you do that; whether you need a regulator for that, whether the regional school's commissioners can do that is I think to be debated. But from my point of view, what you need is to have really strong local partnerships that enable a peer review process to be in place and then somebody who has got an umbrella overview of that, which could well be the regional schools commissioner or it could be a very stripped down form of Ofsted who are only needing to go in and support where the peer review process exposes fundamental weaknesses in schooling at individual schools. (Ned)

Kelvin and Gordon argued that a 'national central body like Ofsted is also necessary' but it should enact area inspection instead of individual school inspection, by which inspectors 'inspect the quality of education for all students in the area at the same time' (Gordon). This national inspection body would produce some comparable standards overall in England to

compare standards between areas and provide 'some national overview of the system to see how it's getting on' (Gordon), with schools visited for that purpose. In other words, the role of that national body should change from 'finding out what they think is wrong' at schools, as has been the case since Ofsted was established, to 'help[ing] the school to improve' and doing this together. According to Gordon, if Ofsted were to be this national body, then this would require a major shift in the relationships between the teaching profession and Ofsted inspectors to working as 'colleagues.' Similar to the Scottish inspection system, that relation should establish a 'partnership' model between inspectors and teachers.

'Area-based inspection' (Kelvin and Gordon) would help to overcome three important problems in the current English education system. First, inspecting 'areas' rather than individual schools allows schools which serve in the same area to work together by avoiding some schools' unfairly gaining advantage over their 'rivals'. Schools working together contributes to system-level improvement by overcoming the competition culture (Gordon, Kelvin). Second, area-based inspection would 'mitigate' the differences in quality between schools in the same area and ensure that receiving high-quality education does not become 'a matter of luck as to where you live' (Kelvin). Third, it would overcome some 'sharp practices' stemming from inspection of individual schools such as student exclusions. Gordon described an alternative approach based on a combination of 'horizontal' and 'vertical' accountability:

> With vertical accountability, we mean something like Ofsted where people from above come in and tell you what to do or tell you what's wrong. With horizontal accountability, this is self-evaluation, this is peer evaluation. Other headteachers from other schools coming in and helping you in a very rigorous way to improve, telling you what's right and what's wrong… You start with self-evaluation. Then move out to peer groups; headteachers in the locality looking to help each other to improve. And then, you get a national body like Ofsted coming in and saying: 'Well, hold on a minute.' The northeast might be too parochial. It might be too inward looking. 'We want to see what your standards are like in comparison to those in London or Birmingham or Manchester'. So, there must be some national overview of the system to see how it's getting on. Some central body making sure that standards overall in England were reasonably comparable and able to look at for example 'What's the quality of mathematics teaching throughout the whole country?' Where good ideas are brought together and put in a report and sent out to all the mathematics teachers. I see a role for Ofsted for doing that kind of work.

> But at the same time, I also think that we need to have local inspectors, not just national inspectors. Local inspectors who are more likely to know the context within a particular school, who know the teachers on a regular basis, who are in and out of the school much more often. And we would be able to see what the strengths of the school are and its weaknesses. Because they're there on a regular basis. You cannot have that in a national system. I mean, there're 25,000 schools in England. You cannot be an expert in all of them simultaneously with a force of 1,500 inspectors... But we have got rid of them because of austerity. They've got rid of the local inspectors. They could be brought back in. To answer your question, in an ideal system we could bring them back in.

> (…) One of the things that I find missing from Ofsted reports is any kind of comparative dimension. They never look at a system elsewhere… Finland doesn't have any inspection at all. They don't have any inspectors. And yet, they turn out the best results. Surprise, surprise. If I was working for Ofsted, I'd think 'How in heaven's name is that possible?' Right. 'The reason is that they invest heavily in their teachers?'… I can see Scotland still does it in a better way. For example, in Scotland, it's called Education Scotland. If you look at the HMI system up there, they begin by saying: 'We do this in partnership'. An inspection must be a partnership with the teachers. And they consider themselves as colleagues, professional colleagues who come to help the school to improve. Ofsted doesn't behave like that. Ofsted comes in to find out what they think is wrong. So, what the teachers in England say is: 'Inspection is done to us. It is not done with us.' I would be in favour of a move towards a partnership between the teaching profession and Ofsted to do this together. (Gordon)

For Cooper, a national inspectorate should take a role only if schools 'go off of the rails too far'. National inspection standards should be used as the 'benchmark' instead of their current use to hold schools accountable and to drive school improvement (Section 7.4.3).

Participants' views on establishing a 'better inspection system' also extended to improvements to Ofsted's methodology and a focus on 'developmental' forms of inspection. The main methodological suggestion pertained to a stronger emphasis on qualitative data to enable inspectors to understand schools in their context. An inspection of a school should be conducted as a 'case study' instead of an 'outcomes-based external quantitative evaluation' (Felicia, Dennis). For example, Felicia noted that 'the qualitative stories behind the quantitative data' explain 'far more' than quantitative data about 'what happened' and what 'doesn't look like a success story on the quantitative data' but actually 'success stories. Dennis, a retired HMI inspector, noted the importance of class visits instead of making student performance data the sole determinant of quality. Dennis also argued for 're-setting the relationship between Ofsted and the teaching profession on the basis of a two-way educational conversation with schools and teachers.' His emphasis on 'conversation' was related to the current problematic 'nature' of Ofsted judgement. Importantly, he emphasised that there is no right way to inspect schools and any inspection findings are 'subjective, contestable, and never definitive'. Inspection is just a 'snapshot', hence the time-specific nature of inspection judgements and Ofsted's inability to comment meaningfully on 'progress'. As he observed:

> in the current atmosphere, inspection is judgemental, arbitrary and a fault-finding accounting process. However, it needs to be developmental, and seen as educational enterprise and principled. In this regard, re-setting the relationship between Ofsted and the teaching profession on the basis of a two-way educational conversation with schools and teachers is very important. In doing so, both inspectors and teaching professionals have to readjust their mindset… (Dennis)

Whilst the majority of participants argued for reframing the current system of school inspections, Orion defended the importance of a national inspection body within the current 'fragmented' school system. As he explained:

> people don't much like being evaluated and as likely as not, they will take some evasive action, or they will engage in some presentational aspect. So, you know, most evaluative systems have unintended or undesirable consequences. That does not per say outweigh the positive benefits. But it is a balance between the negative impact of evaluation and the positive benefits that can be gained from it… once you come to the decision that school evaluation and inspection are a necessary evil; then, the important question is how they can be done in a way which is developmental and genuinely improves the quality of provision for young people.
>
> (…) although I was initially very critical of Ofsted and the way it worked 'which is that it had these kinds of private contractors that it contracted in to train them maybe and they had manuals and so on and so forth.' To be an Ofsted inspector wasn't a full-time job. And I thought that kind of Ofsted processes and practices were highly suspect because of this contracting in of supposedly experienced people and because of the negative impact that Ofsted inspections paid to have on school and teacher morale. And I became too concerned about the independence of Ofsted. I thought, it was too much a creature of government and I still have concerns about that.
>
> But I do think that the existence of an independent or quasi-independent inspector is important and what we have is Ofsted for better or worse… I think, were we to lose Ofsted, we would suddenly realise what an important job of work it in fact does… I do think it's important that we have the capacity to identify failing institutions. What I think is wrong is that we can't apportion blame correctly. So, it's the school that gets blamed. And then, we are, you know, parachuting new headteachers and we find them in a multi academy trust or whatever we put it on special measures and so on and so forth. And the political and contextual variables that make a contribution to the success or failure of a school are not part of this school report. They're not part of the story that's told about the school. That's worrying because it's a partial story.
>
> But I do think, it's important that there is some national body that has responsibility for looking at the quality of education and can look across contexts and local authority and different forms of provision. This is even more important as the system fragments. So, where all schools controlled by local authorities, it would be less important than it is in a context where some schools of local authority controlled, and others are Academy controlled, and some are directly funded by the Department of Education. And there are a lot of independent providers. (Orion)

Orion was, however, against 'routine inspection' because it means that resources are spread 'thinly'. Schools could be classified as 'green', 'amber' and 'red' based on quantitative, contextualised indicators, with inspectors visiting 'red' and 'amber' schools for 'tough inspections.' For 'green' schools, the focus would be on more 'developmental forms of evaluation'. Another alternative would be 'doing less inspection but randomly', although this approach would go against the principles of 'fairness and reasonableness or inclusiveness'. As Orion put it:

My preference would be that we stopped inspecting all schools and only inspected those schools that had a profile of warning lights that gave sufficient cause for concern. One of those warning lights but certainly not the only one might be a pattern of results. But others should be about complaints and absenteeism and expulsion and the like. So, school exclusions will come to harm on my list.

(…) some forms of inspection need to be tough. So, we need inspections that can tell us the truth about what's going on in an institution whether that's a prison or a hospital or a school you know in a public institution. We need to know what goes on in the public's name. Why should schools be different?

However, even though we do need to be able to carefully scrutinise what goes on in schools, the idea that we need to do this to all schools all time as it were or on a rolling programme seems to me to be wrong. There has to be a way of combining evaluation for development and evaluation for tough accountability and one way of doing that I think is to have the tough accountability by exception. So, it's only used when certain kinds of triggers are evident. So, we might say: 'The schools should be inspected when their exclusions reach a certain level. Schools should be inspected when complaints reach a certain level. Schools should be inspected when the performance of young people is significantly below where we would expect it to be, taking into account contextual variables. Schools should be inspected when… you carry on like that'. So, instead of inspecting all schools, instead of subjecting all teachers to this, you do it by exception. You say, in order for a school to be inspected, there needs to have this number of red lights or this number of amber lights. But if it's got a lot of green, we're not going anywhere near you. And then, you need more developmental forms of evaluation.

So, you want some externality and some independence, you want evaluation to be developmental and you want inspection to be protective of the rights of the child as it were, you want inspection to stand up for the quality of education for the child or the relative like wellbeing, safety, access to high quality education and to only take place when those alarm bells go; but, how you do that depends very much on the particular political and organisational context for education.

But the other solution is 'do less of it but randomly'…  I think, it's justifiable on methodological grounds but not on fairness and reasonableness or inclusiveness. Because it would be very difficult to do and be able to talk to a range of people that you would want to talk to. So, you need to give them a fair warning as it were. Well, I don't mean to teachers are. But I do mean you know parents and other stakeholders.

(…) The external evaluation should take account of the full range of quantitative evidence available on the school and should collect direct evidence of contextual variables, school processes and the milieu of the school, its atmosphere and the sociability and things of that kind. So, both qualitative and quantitative in both instances. My own view is that of the external evaluation front, it should pay attention to stakeholders, and it should pay particular attention to the political and contextual variables that are important in understanding why the school performs as it does. (Orion)

## 7.5 Conclusion

This chapter focused on findings pertaining to research question 2, on alternatives to the official approach to school evaluation in England suggested by expert participants. Overall, two key

positions emerged from the data: pragmatic and radical. Pragmatic suggestions were based on incremental changes and 'tweaks' to the current system rather than a radical departure from the established direction of travel. Four main suggestions included: an improved cycle of inspection of 'all' schools including MATs; increasing resources for more comprehensive Ofsted inspection visits; more detailed data on the views of parents and students to add to the current Ofsted data, and more effective support for professional development in schools. Importantly, the support for schools was referred to as the remit of the DfE because the main function of Ofsted is to 'diagnose' and 'inform' (Neil). This does not mean that Ofsted is not to support schools, but its role is more about setting expectations through the Ofsted framework and inspectors' feedback, in order to steer the system. Resources (money and time) were seen as the main reasons why Ofsted cannot visit schools more frequently. Although some participants noted 'political interference' and resources as two key barriers to improving both policymaking and Ofsted inspections (Bella, Eduardo, Neil), no suggestions were cited for addressing these issues.

The alternatives from a radical position, questioned the very *pillars* of the official approach to school evaluation, in order to address two main problems with this approach as currently designed and put into practice. Firstly, problems with the evaluation of the system of public education itself. Secondly, and linked to this, accountability, marketisation, league tables and Ofsted inspection set up in the current system as 'external drivers' of school improvement. The current system was described as focused on 'a single school improvement' (Cooper, Orion), creating winners and losers due to the reliance on competition. Suggestions for re-conceptualising the *pillars* of the current official approach were aimed at school evaluation which would serve 'whole system' improvement. From this standpoint, suggestions included: reconceptualising school evaluation within the 'collective' and 'collaborative' rather than 'competitive' approach to school improvement; priority being given to community and 'voice' over 'choice'; reconfiguring 'accountability' within a complex, contextualised understanding of school improvement; and reframing the system of school inspections. Although the experts' views seemed to converge on collaboration as a foundation for an alternative approach to school evaluation and school improvement, this is where the convergence (commonality of views) seemed to stop. Each of these experts offered different methods and methodologies from this foundation, detailed in Section 7.4.1 above.

Having presented polarised views on how to improve school evaluation in England in the

present chapter, in Chapter 8, I will summarise and discuss the key findings by focusing on points of commonality between the pragmatic and radical stances under the following four themes: schools need support; Ofsted inspections would benefit from improvement; accountability should be 'reconfigured' and school evaluation itself could be 'reconceptualised'.

# CHAPTER 8: DISCUSSION

## 8.1 Introduction

This chapter summarises the key findings in relation to research questions 1 and 2 and, to address research question 3, discusses them in the light of theoretical literature presented in Chapter 3. The discussion is informed by Christie and Alkin's (2013) Evaluation Tree Theory, in particular their conceptualisation of *foundational ideas* underpinning the different branches of the evaluation 'tree', Patton's (2006, 2011, 2012) distinction between traditional and developmental evaluation, as well as arguments about democratic evaluation (MacDonald, 1974, 1976; Norris, 2015; Simons, 1987). As explained in Chapter 2, since 1988 school evaluation in England has become increasingly centralised by successive governments to focus on what I refer to in my thesis as the official approach. Driven by the principles of New Public Management (NPM) and marketisation, the official approach rests on four *pillars*: accountability, Ofsted inspection, school league tables, and parental choice, aiming at the *individual* (single) school improvement. However, researchers have reported a number of negative consequences of this approach (see Chapter 2) and, similarly, experts who participated in this study put forth a number of suggestions for 'tweaking' (Eduardo) or even reconfiguring the official approach and recasting its *pillars*.

Section 8.2 presents a summary of key findings to research question 1 related to experts' views on the official approach to school evaluation in England since 1988. Overall, interview data pointed to a strong polarisation in their views, whereby the same *pillar* of the official approach was assessed either as problematic or a strength in the evaluation of schools. The polarisation appeared to be aligned to Patton's (2006, 2011, 2012) distinction between traditional and developmental evaluation, as well as the underpinning view of school improvement as either simple and linear or complex. Section 8.3 presents the key findings to research question 2 by summarising alternatives to the current official approach suggested by educational experts. These alternatives are discussed in the light of debates in the field of evaluation (see Chapter 3).

Overall, the themes which emerged from interview data highlight, firstly, that schools need support and one of the main objectives of school evaluation should be to identify this support. Secondly, improving school evaluation entails improving Ofsted inspections, through minor changes to the methods of inspection, as well as a 'reframing' of the role of Ofsted. Thirdly,

some experts put forth arguments for 'reconfiguring' accountability and 'reconceptualising' school evaluation. Based on these more radical alternatives, my suggestions for 'recasting' the current *pillars* of school evaluation are discussed in Section 8.4.

## 8.2 The official approach to school evaluation: benefits and problems

Interview data on the official approach to school evaluation revealed a strong polarisation of experts' views, whereby the same *pillar* of the official approach was assessed either as problematic or a strength in the evaluation of schools. In other words, the benefits and problems cited by participants were broadly aligned to their predominantly positive or predominantly critical views. Where the official approach was seen as beneficial, it was because it was considered to be fit for the purpose of driving improvement at the school and national levels. The benefits included: having national standards and frameworks; accountability to Ofsted; high stakes as a tool for 'bringing people into line' (Kent), and transparency to parents by providing them with the 'right information' (Neil, Eduardo, Bella) in Ofsted reports and schools' league tables.

Experts who assessed the official approach as problematic expressed their concerns about its key focus on the *individual* (single) school improvement, instead of *system-level* improvement. The problems cited by these participants included issues with how and why national standards have been used (as prescriptive, rigid, externally set and deployed as a tool of political control); how accountability has been conceptualised (as external accountability 'from above') and enacted (to sanction and punish schools which did not deliver the expected standards), as well as the very design of evaluation around competition in public education, underpinned by a belief in the 'false God' of marketisation (Kelvin). Whereas the same *pillar* was assessed either as a source of strength or problems, this was not associated with the current roles of the experts, the types of their educational expertise, or their gender. This polarisation is aligned to the differences between the traditional and developmental approaches to evaluation (Patton, 2006, 2011, 2012). There appeared to be also an underlying espousal of bureaucratic or democratic values (MacDonald, 1974, 1976). The key factual and conceptual findings (Trafford and Leshem, 2008: 133) from expert interviews are summarised in Table 8.1 and discussed in more detail below.

| Main findings in relation to: | Key benefits cited by experts | Key problems cited by experts | Conceptualising factual findings in light of the literature on evaluation and the conceptual framework of this thesis |
| --- | --- | --- | --- |
| National standards and frameworks | 'clarify' how quality looks like and what is expected<br><br>'make people act' in the 'right' direction<br><br>ensure 'reliable', 'valid', 'fair', 'objective' and 'consistent' judgements (Bella, Richard, Kent, Neil, Eduardo) | set 'rigid' expectations without capturing contextual differences between schools<br><br>prescribe what 'quality' looks like<br><br>have been used to steer the system without much consultation with schools<br><br>their use has decreased innovation and creativity, deprofessionalising teachers<br><br>have been used to judge schools (Cooper, Ned, Torr, Dennis, Gordon, Kelvin, Nora, Felicia, Orion) | Views of experts who expressed predominantly positive assessments were aligned to the traditional approach to school evaluation and an underlying understanding of schools as simple, linear, stable, and fixed (Patton, 2011).<br><br>Views of experts who expressed predominantly negative assessments were aligned to developmental evaluation (Patton, 2011), based on the assumption that the complexity of schools cannot be captured through predetermined indicators in national standards and frameworks. The 'bureaucratic' orientation in the official approach, based on the values of 'contractor' (MacDonald, 1974, 1976), i.e., the Government and Ofsted, was problematised. |
| Accountability | accountability to Ofsted is a tool for 'bringing people into line' (Bella, Kent, Neil, Eduardo) | problems with punitive mechanisms of its implementation<br><br>accountability to Ofsted and government rather than to children, parents, and communities<br><br>externally set accountability is 'counterproductive'<br><br>outcomes-based accountability creates 'perverse' incentives (Cooper, Ned, Torr, Dennis, Gordon, Kelvin, Nora, Felicia, Orion) | Views of experts who expressed predominantly positive were aligned to the 'enforcement' connotation of accountability (Schedler, 1999).<br><br>Experts who expressed predominantly critical views problematised both the assumption that accountability improves schools and its enactment as 'enforcement' (Schedler, 1999). In alignment with Christie and Alkin (2013), they framed accountability as 'being answerable'. Their orientation to evaluation was aligned toto democratic values in which schools have 'voice' (MacDonald, 1974, 1976). |

| Parents as 'customers' in a marketised education system | acts as a 'lever' of school improvement<br><br>transparency is achieved by providing parents with the 'right information', i.e., Ofsted reports and school league tables (Bella, Kent, Neil, Eduardo, Richard) | parental choice was constrained through resources and used for political purposes to attract middle-class votes<br><br>applying market ideas of consumer choice and competition in public education created an 'aggressive war' between schools<br><br>marketisation as a 'false God' and 'recipe' for individualism, lack of coherence, and 'poor' governance of education (Cooper, Ned, Torr, Dennis, Gordon, Kelvin, Nora, Felicia, Orion) | Participants who expressed predominantly positive views were aligned to the traditional approach to school evaluation, with parents seen as service users. Their concern about the involvement of parents as 'outsiders' resonates with bureaucratic evaluation's focus on the contractor's values and autocratic evaluation's focus on 'evidence' (MacDonald, 1974, 1976).<br><br>Participants who expressed predominantly critical views saw parents as 'insiders' and advocated for the inclusion of parental 'voice'. |

**Table 8.1:** Summary of key benefits and problems with the official approach to school evaluation cited by the participants

### 8.2.1 National standards and frameworks 'make people act'

Whilst all educational experts who participated in this study referred to having national standards as essential or understandable, their views on how and why to use national standards differed. As captured in Eduardo's quote in the heading of this section, some participants referred to standards and frameworks as beneficial because they made schools act. Schools improve in a 'mechanism' that sets expectations and provides a 'steer' for people to 'act' in a 'right' direction. Because national standards and frameworks 'clarify' what outstanding schools should be doing, they also decrease differences in the quality of educational provision between schools. Furthermore, their use is beneficial because they ensure 'reliable', 'valid', 'fair', 'objective' and 'consistent' judgements. For instance, there was a consensus that the autonomy of Ofsted inspectors relied on such judgements. Bella and Richard called inspectors 'objective professionals', whose roles are to go to schools to decide objectively on 'how well you are doing' (Bella) based on the national standards in the Ofsted framework. Neil also saw inspectors following 'a standard methodology' in inspection as beneficial. He pointed to training for inspectors in 'a standard methodology' from September 2019 which sought consistency to ensure 'validity and reliability'. Eduardo highlighted the 'expertise of inspectors' which relied on 'some element of standardisation' in order to reach 'fair' and

'reliable' judgements. He thus argued for autonomy for school inspectors, to 'a certain extent', i.e., 'within a framework' designed to 'steer what people do'. As he put it:

> you need to be able to adapt what you do, to what you are finding in the school you are at. So, in that, the expertise of the inspectors is important. Because schools differ. So, there isn't like necessarily an approach you could just mechanically use all the time. You need to be able to follow up avenues. If you find something that seems particularly interesting, particularly problematic, you need to be able to follow that route and not have to stick to a script that can make you lose sight of some really important stuff. Saying that of course, you do need some element of standardisation. Because there's a need for schools to be judged fairly and reliably. So, what you need is 'autonomy within a framework that steers what people do'. (Eduardo)

Unlike Picciotti (2015), who proposed fee-independency to achieve democratic evaluation (see Chapter 3), Eduardo noted that breaking free from government funding might make inspection depend on other sources of funding, which would be problematic:

> The question is then if we were not dependent on the Government for funding who would fund us? So, obviously, the potential is that to get schools to fund you directly to pay fees to be inspected. I think, this runs the real risk of producing capture, because schools are more likely to want to pay for inspection that is going to not be too harsh on them. So, I think that's potentially a problem with that. (…) When you look at the reality, you will see that all inspectors internationally are funded by their government in some way ahead and that's probably a reason for that. Also, because the Government is of course a key stakeholder in getting the information about the quality of the education system. (Eduardo)

The experts who expressed predominantly positive views appeared to hold a linear understanding of school improvement through national standards, which set goals and predetermined outcomes to make people act on them. Their views were aligned with traditional evaluation, which assumes that schools are stable, linear, and fixed (Patton, 2011). The predetermined outcomes and indicators set by Ofsted and the Government fixed the system 'upfront' in search for 'certainty' and generalised evaluation findings to disseminate (Patton, 2006: 30; Patton, 2011: 23-26). In traditional programme evaluation, this leads to employing a 'rigid, mechanical' approach to evaluation (Guijt et al., 2012: 1). The references to 'steering' people in the 'right' direction and objective evaluation through national standards also resonate with elements of bureaucratic and autocratic evaluation (MacDonald, 1974, 1976).

By contrast, experts who expressed predominantly negative views approached national standards with caution and explained their reservations about their current use, citing two problems (see Table 8.1 above). The first problem was *how* they were used, based on setting 'rigid' expectations for schools to comply with, but without considering the financial, social, and political contexts within which schools operate. They were 'too prescriptive' about what

quality should look like in schools and did not include schools' views. The second problem was related to the *why* question, and in particular seeing national standards as a 'political tool' used to steer people, in order to maintain social and political control. Decontextualised, prescriptive standards and frameworks, rigidly used to judge schools as a 'yardstick' of improvement, diminished innovation and the creative capacity of teachers. They deprofessionalised teachers and reduced the complexity of schools.

And yet, for these experts, schools are complex and provide 'phenomenal things' (Felicia, Nora) for students which cannot be captured through standardised predetermined goals, indicators and outcomes. What is more, not fitting into national standards is not a sign of 'failure', because every school is different and maybe 'adapting' to 'emerging' changes in its environment or may be stepping out of the 'described standards' for more 'innovative' and creative work (Patton, 2011, 2012). What counts as a failure in relation to rigid national standards may be a 'seed' of new development and innovation (Guijt et al., 2012: vi; Patton, 2012). Experts who expressed predominantly critical views thus appeared to espouse the developmental approach (Patton, 2006, 2011, 2012) to school evaluation. They also problematised what MacDonald (1974, 1976) refers to as a bureaucratic orientation in the official approach, with national standards reflecting the values of the external contractor (the Government, Ofsted) rather than the values of the school and the local community.

However, as Gordon and Torr pointed out, not only schools' creativity and innovation are constrained by rigid frameworks, but inspectors' autonomy is also 'constrained' within 'compliance culture'. Gordon's example illustrated this as follows:

> For example, a head writing to me who has 57 different languages being spoken in his school. When the inspectorate arrived and looked at the classes of English, they said: 'I'm sorry we're not prepared to take into consideration the fact that there's 57 languages being spoken'. They've either got to pass the tests or not. Now compare that with schools in the North of England, where there's only one percent of the population are from ethnic minorities. Only one percent is very different than London. But the Ofsted inspector said: 'we are not allowed to take context into consideration here. We just look at the results. And the results are poor'. (Gordon)

Overall, experts' views on national standards and frameworks were polarised and aligned to Patton's (2006, 2011) distinction between traditional and developmental evaluation, and schools' being seen either as linear or complex organisations to evaluate. Objectivity was cited as an important benefit by experts who articulated predominantly positive views, despite the growing consensus in the field of evaluation that objectivity is a human achievement and there is always an 'I' in the achievement of objectivity. For instance, aiming for an objective

conclusion ('judgement') is an 'illusion' because subjectivity is always inherent in the 'interpretation' of any type of data, even in the statistical analysis of quantitative data (Berger and Berry, 1988: 7).

### 8.2.2 Accountability: 'keeping people on their toes'?

All educational experts who participated in this research considered accountability to be an essential element in education, but their views about how accountability was conceptualised and enacted differed. Accountability was critiqued in terms of four problems (see Table 8.1 above). First, on the basis of a punitive nature of its enactment and not allowing schools to fully justify their unexpected or less successful outcomes. Second, conceptual problems were cited, whereby accountability was conceptualised as enabling parental choice and driving the *individual* (single) school improvement rather than improving the whole system of public education. Third, equating accountability with inspection was seen as problematic because schools are accountable to a broader audience, from students to the community. The idea of improving schools through accountability was seen as based on an assumption that schools need external drivers to improve. For these experts, accountability is not about improvement, because improving the quality of a school's provision is an internal issue of the school professionals. Rather, the purpose of accountability is for schools, as public institutions, to account to students, parents, governmental agencies and the wider society. The fourth problem with accountability was its outcomes-based orientation, which concentrates mainly on externally set quantitative student performance targets, narrowing broader educational purposes and creating 'perverse' incentives for schools.

The 'enforcement' connotation (Schedler, 1999) of accountability as it is currently used in England was implicit in experts' arguments about the negative consequences of the system, from 'gaming' to school staff retention and staff recruitment crises. This finding resonates with Christie and Alkin's (2013) point that accountability and evaluation overlap. Evaluation may serve accountability by providing information in relation to 'being answerable', but the 'enforcement' aspect of accountability (in response to deficiencies) is not part of the evaluation. When the emphasis in school evaluation is on the enforcement element of accountability, schools will not fully benefit from being evaluated. Participants' understanding of accountability was aligned with Patton's (2012) conception of complexity-based accountability, in which schools should document openly what they struggle with, as well as with democratic values in evaluation, which ensure that those who are being evaluated and the

wider community should have the 'voice' (MacDonald, 1974, 1976).

Experts who held predominantly positive views focused on the benefits of accountability and 'high stakes' accountability as drivers of school improvement, with accountability to Ofsted cited as the main driver of improvement. Because Ofsted inspectors make it clear that schools will be held to account, this 'keeps them on their toes' (Neil) and makes 'people who are doing good' continue to do 'the right things' (Bella). Although the participants noted some potential negative consequences of high stakes accountability, such as decreasing the quality of information collected, they considered high stakes accountability to be necessary as a motivator in 'pushing' people forward. The 'fear' in relation to high stakes accountability was explained as an individual problem that was to be managed by the individuals themselves. These explanations are in line with the traditional understanding of accountability, whereby accountability is employed to 'control and locate blame for failures' by engendering 'fear of failure' through evaluation (Patton, 2006: 30). The participants also highlighted that outcomes-based accountability may be slightly problematic, but it is a useful form of accountability, because school league tables are an important 'end product' of what schools are delivering in exchange for the public money invested in public education (Eduardo, Neil, Richard). Therefore, outcomes-based accountability was described as 'reality check' (Eduardo). Experts who held predominantly positive views referred to two 'traditional' purposes of evaluation: accountability and improvement (Patton, 2006, 2011). Their views were aligned to what Christie and Alkin (2013) term 'social accountability', in which schools are accountable to external contractors. This understanding, however, appeared to conflate accountability as one of the *pillars* of school evaluation, aimed to contribute to school improvement, with accountability as the purpose of evaluation.

### 8.2.3 Competition and parents within a marketised education system

Whilst all experts who participated to this study valued parental involvement, the difference was in viewing parents either as 'outsiders' or 'insiders' in the education system. A marketised education system was not mentioned by participants who held predominantly positive views. However, competition and marketisation were cited as fundamental problems by educational experts who expressed predominantly critical views. These experts problematised parental choice within a marketised education system in terms of three main problems (Table 8.1). First, framing parental choice as a right of parents was a problem because parental choice is restricted to the resources that parents have, such as time, money, networking and connections, as well

as knowledge about how to make school choices. Parental choice was thus interpreted as a tool to win 'votes' from middle-class parents rather than genuine choice (Kelvin, Nora). Second, applying market ideas of consumer choice and competition in public education was flawed and created an 'aggressive war' between schools (Kelvin), with winners and losers. Third, a marketised education system was a recipe for individualism by relying on individual school improvement, rather than system improvement. It also led to a lack of coherence in the system, making it 'fragmented' and suffering from 'poor' governance (Dennis, Gordon, Orion). Market-based strategies were cited as the reason why it is very difficult to plan a coherent system for education in England. These views diverge from claims in the literature that parental choice is a way of enacting democratic education (Box et al., 2001), because for them 'voice' was essential in democratic relations. For instance, the policy which removed parental involvement from Multi Academy Trusts (MATs) was cited as 'very anti-democratic' by Gordon (Chapter 7, Section 7.4.2). The participants appeared to hold the view that the emphasis on parental choice over voice makes parents 'outsiders' in the school system (see also Section 8.3.4).

As mentioned above, experts who held predominantly positive views mostly kept silent about competition and marketised education, except for Eduardo. His interpretation of the nature of competition with both benefits and negative consequences highlights questions about equity:

> you actually look at the impact of competition in education what you see is typically that there is a positive, though not a particularly large effect, on outcomes. So, I think, in that sense, competition works. But there's also a negative, but again not very large effect on equity. So, you get no full improvement, you also get a greater distinction between pupils from different social backgrounds as well. So, that then really leaves you with the policy question of A) which of those things do you want to put to the forefront and B) do you think you can mitigate that problem of equity in other ways? (Eduardo)

These experts emphasised the legislative right of parents to choose a school, in response to my probe on their views on parental choice and a marketised education system. Parental choice was cited as a driver of school improvement because parents hold schools to account through their school choice. Parental choice thus acts as a 'leverage' on schools to improve so that they can be chosen by parents for their children. The role of school evaluation was cited as helping parents to make an informed choice with a reliable, easily understood source of information in the form of Ofsted reports. These views were aligned to the traditional approach to school evaluation, with parents seen as the users of evaluation findings. However, as Patton (2012) suggested, using evaluation findings to make practitioners accountable is not a route to

developing high-performance organisations, as these rely more on development than accountability. On the one hand, the emphasis on understandable and reliable sources of information, as well as the importance of parents' views to inform inspection visits, shows elements of MacDonald's democratic evaluation through 'accessibility' of reports for 'non-technical audiences' and 'inclusiveness' of wider stakeholders (MacDonald, 1974, 1976). On the other hand, experts who expressed predominantly positive views did not foreground the value of the whole community in the evaluation process. They referred to wider parental involvement only when prompted. For instance, Bella and Neil explained why wider parental involvement is problematic, based on the premise of autocratic evaluation (MacDonald, 1974, 1976) that evaluation is for 'professionals' (Bella) and that there is no evidence in the literature that parental involvement is of benefit in schools' decision making (Neil). Neil, Richard and Eduardo highlighted the potential conflict in the values of parents who may have extremist views which do not fit into an 'agreed set of values that run through our country and [are] reflected in legislation' (Neil). As Neil put it:

> The principle you've got to ask yourself is about the purpose for inspection and your theory of action. You've got to ask yourself about your evidence base. So, there isn't evidence I don't think that parental involvement in shaping the curriculum and in determining decisions made in the school leads to higher quality education. I don't think, there's any evidence that says: 'that is the case'. Or certainly that there's no consensus in the academic literature willing to this. So, we don't think it's our job to get into encouraging schools to make that sort of contact happened.

> And we have an increasingly polarised and fragmented civic life in England at the moment. And increasingly agitating groups are using parents as a way to try and change the curriculum within a school because it doesn't agree with their values. So, that includes vegans, and it includes some Muslim groups, and it includes other religious groups as well. And there're Christian groups in particular places, Jewish groups in other places, Sikh groups are in other places. And that undue influence of parents is something that as a society and politically we've chosen that we ought to resist. Because we have an agreed set of values that run through our country and that's reflected in legislation. So, it's a very long-winded way of saying, there are risks to parental involvement and there aren't obvious benefits in terms of them shaping the curriculum.

> But, in inspection, in terms of the evaluation of what's going on, in the experience of my children, I think should have a much bigger place for parents to rely/agree on that. And it's the practicality of inspection length and money that's stopping us from doing that. And if I could change those three things that we do a lot more in that space. (Neil)

Eduardo's emphasis on the 'societal role' of education that 'goes beyond parents' appeared to imply that micro-democratic values might be conflicting with macro-democratic values:

> of course, the views of parents are important. They are key stakeholders in the school. But it is not always an automatic thing to get those of views. I mean, we do a spot-far inspection process when parent surveys, parent view. The response to that is quite

differential... [inaudible]. So, you can have a problem if you're not careful that only the loudest parents get heard. And that does not necessarily represent the body of parents as a whole. So, that would be one issue I can see with getting them to make a decision. Of course, parents do not necessarily all agree with each other either that- so, that's often a problem as well...

Education has a societal role that goes beyond parents and sometimes the views of parents can conflict with what we see as important societal rules of the school. An example of that is the recent issues we have seen in some schools around the education of children around LGBT issues. So, my view is that sometimes obviously you need to have a dialogue with parents. Because some of those issues come from misconceptions about what is actually being taught. But, at the end of the day, as a society, we do have certain broad values and sometimes we have to encourage schools to teach those also when parents do not particularly like them. And you see that with various forms of equalities and protect characteristics happening. We've seen that with LGBT. But we can also see that for example not all parents are happy if schools promote gender equality... And yet as a society, those are important values to us. So, I think we sometimes need to also go a little bit against some parents. (Eduardo)

Overall, each expert who participated to this study appeared to speak about 'democratic values' by pointing to parental involvement. However, what they meant by parental involvement differed in terms of construing parents either as 'outsiders', or 'insiders' in the education system. The participants who conceptualised parents as 'outsiders' referred to the contractor's values that feature in bureaucratic evaluation and to evidence, which is a priority in autocratic evaluation (MacDonald, 1974, 1976). However, those participants who saw parents as 'insiders' prioritised democratic values inclusion of parental 'voice'.

## 8.3 Improving school evaluation in England

The suggested alternatives to the current official approach (research question 2) have been discussed in Chapter 7 in light of what emerged as the radical or pragmatic stance taken by participants. This section presents themes pertaining to the suggested alternatives. These themes highlight that 'schools need support' and the process of school evaluation should identify this support as one of its objectives (Section 8.3.1). Another theme reveals that improving school evaluation would entail improving Ofsted inspections, which could involve minor changes to the methods of inspection, as well as 'reframing' the role of Ofsted (Section 8.3.2). The remaining two themes which emerged from the data include 'reconfiguring' accountability (Section 8.3.3) and 'reconceptualising school evaluation' (Section 8.3.4). Whilst this section seeks to depart from the polarity in experts' views, it is important to explain that a pragmatic stance was associated with 'tweaks' to the current system, underpinned by an awareness of resource limitations (Eduardo, Neil), as well as a belief that 'it's just the way it

is' (Kent) and 'not that we can do anything about it, because that's the system in which we live' (Bella) (see also Chapter 7). The radical stance generated far-reaching alternatives, characterised by 'reconceptualising', 'reframing' and 'reconfiguring' the system, leading to my suggested 'recasting' of the current *pillars* of school evaluation discussed in Section 8.4.

| Themes related to improving school evaluation in England | Factual findings: alternatives suggested by experts | Conceptualising factual findings in the light of the literature on evaluation and conceptual framework developed for this thesis |
| --- | --- | --- |
| Schools need support | support as part of the current approach to school evaluation, based on Ofsted 'diagnosis' but actioned by the DfE and other stakeholders (Eduardo, Neil, Bella) | support as part of the current approach focused on the 'instrumental' use of evaluation findings (Patton, 1996), e.g., by teaching schools how to use evaluation findings |
| | support in the form of a more collaborative approach to school evaluation (Cooper, Ned, Torr, Dennis, Gordon, Kelvin, Nora, Felicia, Orion) | support in the form of a collaborative approach to school evaluation (with new partnership relations between the school and the evaluators) focuses on the process use (Patton, 1996) of evaluation, opening the possibility for schools to engage with the evaluation process itself. |
| Improving Ofsted inspection: from improving inspection methods to reframing inspection | increasing the length and frequency of Ofsted visits and an improved cycle of inspection for 'all' schools, abolishing the proportionate inspection policy of DfE (Eduardo, Neil, Bella, Kent, Richard). | The suggestions to improve Ofsted's methods of inspecting schools were linked to an assumption that the current approach is working for 'all' children because trust-based accountability/inspection would not work for all children (Bella, Neil, Kent, Eduardo). |
| | reframing Ofsted inspections, ranging from area-based inspection, replacing routine inspection with inspection to troubleshoot and inspection based on improved methodology (i.e., partnership-based) (Cooper, Ned, Torr, Dennis, Gordon, Kelvin, Nora, Felicia, Orion) with modifications of inspectors' role, by acting as 'professional adviser' (Cooper) or 'professional colleague' (Gordon) | The suggestions for reframing Ofsted inspections were underpinned by a recognition of the vital role of teaching professionals in school improvement, as well as the importance of trust in effective, honest school evaluation (Cooper, Ned, Torr, Dennis, Gordon, Kelvin, Nora, Felicia, Orion). |

| | | |
|---|---|---|
| Reconfiguring accountability | There was consensus between experts that there was no desirable alternative to accountability. Some participants suggested reconfiguring accountability:<br><br>to embrace a broader educational purpose, such as preparing children for living in society<br><br>introduction of values-based rather than outcomes-based accountability<br><br>avoiding the use of accountability as a means of pressure on schools (Cooper, Ned, Torr, Dennis, Gordon, Kelvin, Nora, Felicia, Orion). | Accountability should not function as a means of pressure and sanctions by relying on 'enforcement' (Schedler, 1999). These views were broadly aligned to the notion of complexity-based accountability (Patton, 2012) in which accountability is used to document both failure and success of schools' efforts to improve. Complexity-based accountability 'questions' people but allows them to 'justify' their actions. At the heart of accountability is 'responsibility'. |
| Reconceptualising school evaluation | All participants valued 'learning exchange' between schools.<br><br>Some participants framed collaborative learning as a systematic enactment of collaborative evaluation in order to improve provision (Cooper, Ned)<br><br>Some participants saw improvement as the only purpose of collaborative learning (Eduardo, Neil, Bella)<br><br>Some experts suggested reconfiguring school evaluation (Ned, Felicia, Gordon, Dennis, Cooper, Kelvin, Orion) | The participants who favoured a systematic enactment of collaborative evaluation argued for the necessity of reconceptualising school evaluation and improvement within the 'collective' and 'collaborative' rather than 'competitive' approach. The underlying reason for this reconceptualisation is to focus on 'whole system' improvement, not 'single school' improvement.<br><br>The suggestions for reconfiguring school evaluation were related to the vital role of peer learning and collaborative evaluation. These suggestions align with developmental evaluation (Patton, 2011) as well as democratic evaluation of schools (MacDonald, 1974, 1976). |

**Table 8.2:** Summary of key alternatives official approach to school evaluation suggested by the participants


### 8.3.1 Schools need support

There was a consensus amongst educational experts who participated in this study that schools need support. Some participants envisaged the support as part of the current approach to school evaluation, based on Ofsted's 'diagnosis' but actioned by the DfE and other stakeholders,

central to which was more training to support schools' professional development. This support was envisaged to function as a 'treatment' of symptoms 'diagnosed' by Ofsted. Other participants envisaged the support in a form of a more collaborative approach to school evaluation, central to which were partnership-based relationships between schools and their evaluators. The evaluators' report should reflect the narrative story of schools, instead of the current judgement of schools based on Ofsted grades and student performance data reported in school league tables.

These findings point to different ways of conceptualising support for schools, with the 'diagnostic function' limited to a simple problem-solution cycle, as well as a broader conception that encompasses supporting schools in the use of evaluation and extends beyond the simple use of evaluation findings to the use of the evaluation process (Patton, 1996). The participants who envisaged support as part of the current approach through more training appeared to focus on the instrumental use of evaluation findings, by teaching schools how to use evaluation findings. The participants who suggested support in a form of a more collaborative approach to school evaluation, within new partnerships between the school and its evaluators, appeared to focus on the process use (Patton, 1996) of evaluation through dialogue between evaluators and schools, opening the possibility for schools to engage in the evaluation process itself.

### 8.3.2 Improving Ofsted inspections: from improving inspection methods to reframing inspections

Suggestions for improving Ofsted inspections ranged from improving Ofsted's methods of inspecting schools to reframing inspections. The participants who saw Ofsted inspection as an essential element in the education system referred to: increasing the comprehensiveness of Ofsted inspection visits by increasing their length and frequency; an improved cycle of inspection for *all* schools and abolishing the proportionate inspection policy of the DfE (see also Chapter 7, Section 7.3.1). For them, not inspecting every school through a regular cycle of inspection caused 'perverse incentives', such as teaching to the test based on student performance data, rather than a more rounded education. Also, they argued that Ofsted should be inspecting MATs, similar to inspections of local authorities.

These could be considered minor changes and experts who suggested them articulated interesting views on 'trust-based accountability'. The defence of inspections was summarised as 'give me an inch, I'll take a mile' (Kent), implying that schools can be trusted only when

184

they are inspected. Neil and Bella talked about accountability through inspection and trust, emphasising that trust and accountability (exercised during inspection visits) are not 'mutually exclusive' (Neil) and not 'black and white' (Bella). Neil pointed to the time in the 1960s, when the Government trusted schools but did not hold them to account, as a situation which worked 'really well for some proportion of the population' and did not serve the 'aspiration for all children', because success depended on whether schools were 'self-sustaining' or not. Bella had 'massive amounts of respect for the vast majority of teachers', but she also saw 'some teachers who are lazy and… need a rocket behind them', hence the need to inspect schools. Therefore, 'identifying' those teachers to 'weed out' and 'close or change [schools] into something else [academies]' is 'essential' because 'there are schools, and there always will be, these are not serving the needs of the pupils properly. Otherwise, it is not 'fair' for students.'

Some participants argued for a reframing of Ofsted inspections, based on area-based inspection, replacing routine inspection with inspection to troubleshoot, and inspection based on improved methodology (i.e., partnership-based inspection) (see Table 8.3 below).

| Area-based inspection | Kelvin and Gordon: A 'national central body like Ofsted' should enact area inspection instead of individual school inspection. |
|---|---|
| Replacing routine inspection with inspection to troubleshoot | Orion: A national inspection body is a 'necessary evil'; however, they should not conduct 'routine inspection'. Instead, 'tough accountability' should be enacted through inspection when schools show 'red' warning lights based on their comprehensive but contextualised outcomes. For schools which have 'green' light, an evaluation would follow the 'developmental evaluation' approach. A random inspection should be used as an addition to risk-based inspection for all schools. |
| | Ned: External evaluation should be providing 'an umbrella overview' when the peer review process exposes some 'fundamental weaknesses.' |
| | Cooper: A national inspectorate should take a role only if schools 'go off of the rails too far'. Performance results of schools or areas should be 'benchmarked' by inspectors instead of their current use to hold schools to account and to drive school improvement. |
| Inspection based on improved methodology | Felicia and Dennis: School's inspection methodology should change, and should work with more qualitative data, considering the context of schools and work with more partnership with schools. |

**Table 8.3:** A range of ideas for reframing school inspections

Experts who put forth the above suggestions for reframing Ofsted inspections also emphasised the vital role of teaching professionals in school improvement, as well as the importance of trust in effective, honest school evaluation (Orion, Ned, Kent, Cooper). As discussed in Chapter

3, there is tension between trusting professionals and the function of inspection as an important element in school evaluation which provides information to governments for the purpose of 'social accountability' (Christie and Alkin, 2013).

### 8.3.3 Reconfiguring accountability

There was consensus amongst experts that there is no alternative to accountability. However, for a more sensitive accountability system, some participants suggested reconfiguring accountability within a complex, contextualised understanding of school improvement. For some experts, there was 'nothing' before 1992, which was 'year zero' in terms of accountability. For others, this was not the case, as schools used to be held to account by Her Majesty Inspectorate (HMI) Inspectors and local authority advisers. This difference might be because there were no 'official judgements of quality' (Ball, 2003: 224) before Ofsted set out national standards, embraced as accountability. According to all participants, the crucial change introduced with the Education Reform Act (ERA) of 1988 was *national* standardisation and *national* accountability.

One idea for reconfiguring accountability was based on going beyond outcomes-based accountability (e.g., student outcomes, attendance rates and test results) to embrace a broader educational purpose, such as preparing children for living in society. Cooper, Ned and Felicia proposed a national debate about 'what schools are for' in order to reconfigure accountability. Another suggestion relied on values-based rather than outcomes-based accountability, stemming from a debate on 'whose values' should be reflected in the standards by which schools are held to account. Contextualising the values and needs of schools and those of the local community was claimed to make a significant difference to schools' responses to the demands of accountability.

However, there was a difference between how experts envisaged operationalising this more complex, contextual accountability. While some experts suggested making schools accountable for nationally defined standards but contextualising them at the school level (Orion, Felicia, Dennis, Gordon), some proposed a locally brokered accountability, designed purely around local values (Cooper and Ned). The proponents of a locally brokered accountability were against using national standards to hold schools to account and favoured creating a unique accountability framework. This unique framework would be based on the values of the school and its local community, with national standards used to inform schools where they are (rather than judge schools) and to inform the inspectorate for benchmarking

purposes. Another suggestion included approaching accountability as a 'shared endeavour' by a wide range of stakeholders, including both the local community, local authorities, and national policymakers. Accountability would thus go beyond 'being purely school based', with all stakeholders 'properly' informed about the strengths and weaknesses of policies that impact schools by a 'properly independent inspectorate'. Another suggestion was that accountability should not function as a means of pressure, sanctions, and reliance on 'enforcement' (Schedler, 1999). Through accountability, people should be able to learn from their 'mistakes', without 'being afraid', and should be given an opportunity to explain and justify the reasons 'when something isn't as expected'. The views of these participants were thus aligned to Patton's (2012) description of complexity-based accountability, in which accountability is used to document both failure and success of schools' efforts to improve. Complexity-based accountability *questions* people but allows them to justify their actions. At the heart of complexity-based accountability lies *responsibility*.

### 8.3.4 Reconceptualising school evaluation

All participants valued 'learning exchange' between schools. However, whilst some participants envisaged collaborative learning as a systematic enactment of collaborative evaluation in order to improve provision (e.g., through school-to-school peer review, Ned, Felicia, Gordon, Dennis, Cooper, Kelvin), others saw improvement as the central purpose of collaborative learning (Neil, Eduardo, Bella). When prompted, these participants noted that collaborative learning is not a substitute for national inspection, because such forms of collaborative evaluation lack national standards and can therefore turn into 'friendly' conversations, with no authority of one school over another. Collaborative learning was also deemed not to be sustainable in the long term, as it requires much more financial resourcing than national inspection.

The participants who favoured a systematic enactment of collaborative evaluation argued for the necessity of reconceptualising school evaluation and improvement within the collective and collaborative rather than competitive approach. The underlying purpose of this reconceptualisation is to focus on *whole-system* improvement rather than *individual* improvement of the 'single' school. The current, 'ill-conceived' system, based on competition and financial mechanisms, caused a 'separation of quality' rather than 'general improvement' at the level of the system. Although these experts' views seemed to converge on collaboration as a foundation for an alternative approach to school evaluation, this is where the convergence

of views seemed to stop, because there was no consensus on how to operationalise the collective and collaborative approaches. Each of these experts then offered a different set of methods, summarised in Table 8.4 below.

| Participant | Proposed methods |
|---|---|
| Kelvin and Gordon | A combination of school self-evaluation, peer review and area-based inspection. |
| Cooper | A combination of school self-evaluation and peer inspection or peer appraisal, with schools that work within the same local authority/local area. National standards should be used to help schools 'benchmark' with others, or one area to another area, and should not be used to 'judge' schools for accountability purposes. |
| Ned | A combination of self-evaluation and school-to-school peer review with schools within the same local authority/area. Ofsted-type of national inspection may not be needed if 'really strong' local partnerships are established, and peer reviews are 'rigorous.' Then, an 'external eye' is necessary to check the sufficiency of the school-to-school evaluation practices, and this role might belong to Ofsted. |
| Felicia and Dennis | A combination of national inspection, peer to peer review and school self-evaluation with schools within the same local authority/area. Ofsted inspection should stay but needs to shift from outcome-based inspection to inspection that takes into account the context and acts in partnership relation. |
| Orion | An evaluation system to support 'incremental improvement' through evaluation is the best way. Schools should be able to learn from their own experiments. Critical friendship between schools helps schools. A combination of school self-evaluation and Ofsted inspection but following the developmental evaluation approach for schools which have 'green' light. 'Tough accountability' for schools which are underperforming and have 'red' light. |

**Table 8.4:** Ideas for reconceptualising school evaluation

Overall, the vehicles for collaborative learning and collaborative evaluation cited by participants included: 'peer review' by schools between schools (Ned, Felicia, Gordon, Dennis), 'peer inspection' (Cooper), 'peer appraisal' (Cooper), 'critical friendship between schools' (Orion) and 'cross-fertilisation' (Kelvin, Ned). These participants saw a place for an independent 'external eye', but each expert had his/her own ideas about how this could be operationalised in practice. This 'external eye' did not necessarily belong to Ofsted. If Ofsted were to retain their role of the schools' inspectorate, then this role should be 'modified', i.e., acting as 'professional adviser', 'professional colleague' and partner. How they inspect should also change, i.e., from a single school to area-based or replacing routine inspection with inspection to troubleshoot, or inspection based on improved methodology (i.e., partnership-based inspection). External evaluation should be focused on managing the risk of failure rather than steering the system according to rigid, standardised criteria. Importantly, the partnership relation was not limited to schools and the evaluator but also extended to schools' collaboration with the broader community. The participants suggested a role for parents and the community

as 'insiders' in the system, in which schools are supported through the involvement of a wider community. These suggestions align with developmental evaluation (Patton, 2006, 2011) as well as the democratic evaluation of schools (MacDonald, 1974, 1976). The involvement of the stakeholders in the evaluation of schools would be predicated on hearing the 'voice' (or voices) of schools, parents, students and the wider community. This approach to collaboration was considered to be a more effective driver of school improvement than marketisation based policies as it encourages collaboration rather than competition, enabling schools and pupils to 'flourish' and be supported by the local community. Within this approach, 'usable' evaluation findings and process use of evaluation (Patton, 1996) would be deployed to support the development of schools rather than use these findings to 'punish' and 'name and shame' schools.

## 8.4 Recasting the pillars of school evaluation in England

This section presents my suggestions for 'recasting' the current *pillars* and *foundations* of school evaluation, based on the views of participants who argued for more radical alternatives to the current official approach (Cooper, Ned, Nora, Gordon, Felicia, Kelvin, Dennis, Orion, Torr). The alternative *pillars* and *foundations* are derived from suggestions put forth by experts for: 'reframing' inspection (Kelvin, Gordon, Orion, Cooper, Ned, Felicia, Dennis); 'reconfiguring' accountability (Kelvin, Gordon, Cooper, Ned, Felicia, Dennis, Orion, Torr), and 'reconceptualising' school evaluation (Kelvin, Gordon, Cooper, Ned, Felicia, Dennis, Orion), together with the theoretical literature by Patton (1994, 1996, 2006, 2011, 2012), MacDonald (1974, 1976) and Christie and Alkin (2013).

In alignment with alternatives to the official approach suggested from a radical stance, this thesis argues that an evaluation system which is more attuned to the realities of schools, students and professionals should be founded on a complexity-based understanding of schools, school improvement, and school evaluation, as well as democratic values (MacDonald, 1974, 1976). On these *foundations*, the *pillars* of school evaluation would be built to include: voice, trusted professionals, capacity building and collaboration. The main purpose of school evaluation would be fostering collaborative learning. Within this approach, improvement would be driven for further future development of schools rather than from assumptions about problems or flaws with schools as they are now. Whilst evaluation would be conducted by considering schools' needs, the governmental need of public accountability would be addressed through national control and inspection, also stemming from the foundation of democratic

189

values and complexity. These alternative *pillars* are presented in Figure 8.1 and discussed in more detail below.
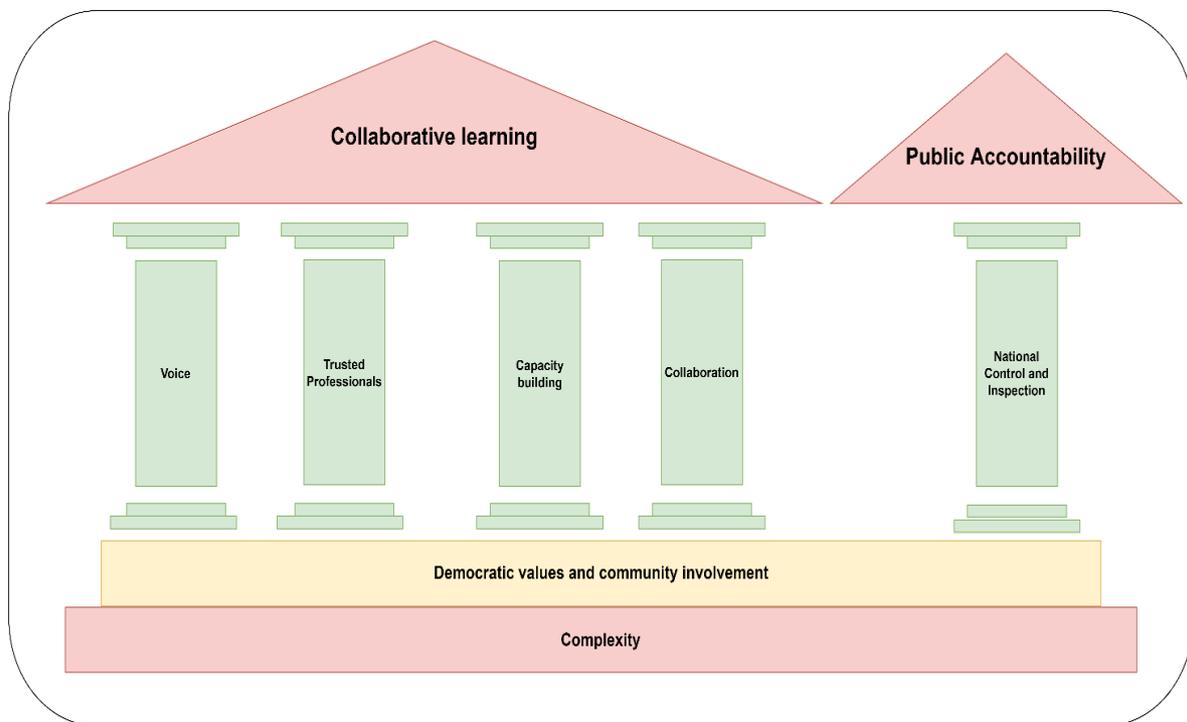


**Figure 8.1:** Alternative pillars of school evaluation

The importance of voice in school evaluation was emphasised by Nora, Kelvin, Gordon, Orion, Cooper, Ned, Felicia and Dennis. According to these participants, voice enables school teachers and leaders to become full partners in the evaluation process aimed at collaborative learning. Whilst collaborative learning could be construed as serving the purpose of school improvement, a more radical perspective recasts collaborative learning as the central purpose of school evaluation, aligning it to the developmental approach (Patton, 1994, 2011). Together with other *pillars*, collaborative learning draws on the *foundational ideas* (Christie and Alkin, 2013) which acknowledge complexity, respect community and rest on democratic values proposed by theorists such as MacDonald (1974, 1976) and Simons (1987). Voice is also connected to the 'insider' role of parents and the local community in the education system, as opposed to the 'outsider' role resting on the *pillar* of 'choice' (see also Chapter 3).

When voices of teachers, parents and the local community are heard, trust can grow between schools and their evaluators (Cooper, Ned, Nora, Gordon, Torr, Kelvin, Felicia, Dennis). Trust and professionalism are reciprocal and trusting professionals who are being evaluated is a

precondition not only of credibility but also the utility of evaluation. When schools are not punished for outcomes that are beyond their control, and when their voices are heard, there is no need for 'fabricating' data and 'gaming' the system. What is more, trusted professionals are likely to be more open to innovating and adopting new practices, seeing risk and failure as a 'seed' of development (Guijt et al., 2012: vi) and, in turn, also being able to develop trust in themselves (Nora). Whilst trust might, from the Government point of view, be related to risk, trust relationships between the evaluators and the evaluands can generate more accurate information for evaluation and can be a more effective basis for schools' ongoing development. By contrast, generating high quality evaluation is extremely difficult in low-trust contexts, especially where high stakes are attached to evaluation (Patton, 2006, 2012). Instead of trying to impose their values on schools, evaluators should work with schools to understand their values and what development and improvement may look like for them. Within this approach, evaluators can still ask 'tough' questions to 'test' school's beliefs as to what they value (Guijt et al., 2012: 18). For this kind of relation, trust plays a vital role.

Whilst much has been published about capacity building in schools for self-evaluation and self-improvement (Blok et al., 2008; Nelson et al., 2015; Nevo, 2009), capacity building can also be recast in terms of *building capacity in evaluation* (Patton, 2015). Expert participants who alluded to building capacity in evaluation included Orion, who argued that schools should be able to learn from 'their own experiments', as well as Dennis, Kelvin and Ned, who pointed to peer-to-peer evaluation as enabling schools to develop more confidence in their own judgements. Evaluative capacity building would also be one of the benefits of developmental 'external' evaluation because the evaluand would also actively take a part in the evaluation, for example contributing to the evaluation questions, designing methods and drawing conclusions. In this way, increased capacity in evaluation would also enhance the 'utilisation of evaluation' (Ryan and Cousins, 2009: 169), with the evaluands developing their evaluation literacy and improving their reflective capacity by learning evaluative thinking. An increased evaluation capacity is essential if schools are to make effective use of evaluation for their 'organisational development and learning' (Ryan and Cousins, 2009: 169) and, importantly, also learn from the very process of evaluating and being evaluated (Patton, 1996). Whilst flexible evaluation designs support the capacity of schools to use evaluation data to adapt to their changing environment, rigidly implemented national standards can be problematic. Similarly, collaboration within and between schools has been considered to be beneficial, but also considered to be constrained by competition and marketisation of education (Ball, 2003). It is

difficult to build a system which improves through collaboration while schools are competing for students and for position in school league tables. Another important aspect of collaboration pertains to collaboration between the evaluators and schools.

In relation to the issue of accountability, it is impossible to deny its importance. Every single expert participant emphasised this, as noted above (Section 8.2.2 and Section 8.3.3). It could, therefore, be argued that public accountability of schools for the provision they deliver to students has been, and needs to remain, an important purpose of evaluation. However, drawing on the developmental rather than traditional evaluation, public accountability would need to be based on the notion of 'answerability' (Schedler, 1999). As discussed above, 'answerability' means that schools are under an obligation to inform and account, in order to explain and justify their decision-making. In this sense, answerability would involve monitoring and oversight rather than rewarding desirable outcomes and punishing outcomes, as has happened in England in cases of 'naming and shaming' underperforming schools (Bates, 2016). As Christie and Alkin (2013) emphasised, 'enforcement' cannot be a part of evaluation, because evaluation serves accountability only in relation to schools providing information for answerability, i.e., being answerable. Based on the arguments developed by democratic evaluation scholars (MacDonald, 1974, 1976; Norris, 2015; Simons, 1987), public accountability of schools would stem from a shared understanding that no evaluation is value free, but all evaluation should serve the interests of children, parents, and society. Public accountability would not be the end in itself but would serve as means to schools becoming more democratic as institutions and offering a truly 'educational' service (Simons, 1987).

It is also impossible to deny that school inspection can be effectively utilised to serve the purpose of public accountability. However, the current role and methods of Ofsted inspections would benefit from recasting or at least modification. From a radical stance taken by some expert participants (Cooper, Ned, Nora, Gordon, Felicia, Kelvin, Dennis, Orion, Torr), the role of a national inspectorate should be recast as ensuring public accountability, thus addressing the basic governmental need to check that the taxpayer's money is well spent by schools, as well as ensuring the safety of students. This could be done either by monitoring schools' collaborative school evaluation or taking a 'risk management' role (gov.uk, 2020).

## 8.5 Conclusion

This chapter summarised the key findings from this research and discussed them in the light of evaluation theory, focusing also on improving school evaluation in England. The key messages

from experts who participated in this study pointed to the importance of support for schools, as well as an argument that the process of school evaluation should identify this support as one of its objectives. Also, improving school evaluation would entail improving Ofsted inspections, which could involve minor changes to the methods of inspection, as well as a more radical 'reframing' the role of Ofsted. Finally, the remaining two themes included 'reconfiguring' accountability and 'reconceptualising' school evaluation. Based on suggestions put forth from a more radical stance, this chapter has also focused on recasting the *pillars* of school evaluation. In alignment with some experts and evaluation theory, an alternative approach to school evaluation would rely on voice, trusted professionals, capacity building, and collaboration. These alternative *pillars* rest on the foundations of democratic values and community involvement, and the evaluation of schools as complex, non-linear organisations that operate in specific contexts. The main purpose of complexity-based evaluation should be collaborative learning because this is essential for improving both the single, individual school, as well as the entire education system.

# CHAPTER 9: CONCLUSION

## 9.1 Introduction

The 1988 Education Reform Act (ERA) became a turning point in education in England, introducing a new 'official' (government driven) approach to school evaluation which rested on four *pillars*: accountability, Ofsted inspections; national testing (high stakes tests) reported in school performance tables (league tables), and parental (customer) choice. The overarching aims of school evaluation were framed as the *individual* school improvement and public accountability. This approach responded to rising concerns about the quality of schools and the effectiveness of local authorities and drew on tools developed within New Public Management (NPM) and marketisation (Norris and Kushner, 2007). The official approach has since evolved but has, to this day, remained controversial, attracting both advocates and critics (Chapter 2). This study has sought to contribute to the ongoing debates, guided by the following research questions:

1. What are educational experts' views on the official approach to school evaluation promoted in England since 1988?

2. Which alternatives to the current official approach have been suggested by research participants?

3. How can these alternatives be assessed in light of knowledge in the field of evaluation?

Expert (elite) interviews were employed to capture the views of educational experts on the official approach to school evaluation and discuss alternatives. Fifteen educational experts were purposively selected because of their diverse professional experience. Their roles included policy development and control as *policy actors*, research and policy critique as *policy influencers*, as well as being at the receiving end of policies and approaches to school evaluation as *school practitioners*. 'Expertise' in this study thus referred to professionals who have historical, technical, theoretical, and practical knowledge of education, as elaborated in Chapter 5. The interview data were analysed for similarities, differences and themes across the participant sample and interpreted through the lens of evaluation theory. In particular, Patton's (1996, 2006, 2011, 2012) distinction between traditional and developmental evaluation, Christie and Alkins' Evaluation Theory Tree (2013), and MacDonald's (1974, 1976) democratic evaluation informed the discussion of key findings from this study.

This chapter summarises the key findings in relation to the research questions (Section 9.2).

Section 9.3 presents the contribution of this study. Section 9.4 reviews the limitations of the study and Section 9.5 suggests topics for future research. Section 9.6 presents my reflection on my research journey.

## 9.2 Summary of key findings

Research question 1 sought to find educational experts' views on the official approach to school evaluation promoted in England since 1988. Participants' views pointed to a strong polarisation, whereby the same *pillar* of the official approach was assessed either as problematic or a strength in the evaluation of schools. Where the official approach was seen as beneficial, it was because it was seen as fit for the purpose of driving improvement at the school and national levels. Where the official approach was assessed as problematic, the main concern was about its focus on the *individual* (single) school improvement, instead of the *whole system* improvement.

Due to this polarisation, national standards and frameworks can be seen as beneficial or problematic. The benefits cited by participants were related to the intention behind standardisation to 'steer' people towards predetermined outcomes and make people 'act' in a 'right' direction, in line with the expectations of the contractor (i.e., government). These benefits were underpinned by an assumption that the 'right' direction can be ascertained through objective, evidence-based inspection. On the other hand, national standards were seen as problematic due to being implemented rigidly, as if every school operated within the same social and economic context. This way of implementing standards was seen to decrease schools' creativity and innovation and undermine teacher professionalism. This study concludes that having national standards is essential, but we need to rethink *how* and *why* they are used.

Another conclusion pertained to accountability, which was considered essential but also in need of a reconfiguration. Whereas accountability is an essential element in education, how it is conceptualised and enacted leads either to problems or benefits. Problems arise when accountability is enacted in a punitive way when schools do not deliver the expected standards. The problems with accountability in the English system as cited by participants included: equating accountability with Ofsted inspection; relying on external accountability based on the values 'from above', and outcomes-based accountability which concentrates mainly on quantitative student performance data, narrowing down broader educational purposes and creating 'perverse' incentives for schools. However, the current conceptualisation and

enactment of accountability can be seen as beneficial, if the underlying assumption is that schools need to be incentivised or 'pushed' through high stakes, within an overarching conceptualisation of accountability to Ofsted. Within this conceptualisation, sanctions and punishments 'bring people into line' and keep them on their 'toes'. Importantly, questions about *whose values* schools are to be accountable for and *to whom* schools should be accountable merit renewed debate and dialogue involving all educational stakeholders.

Whilst experts' views on parental involvement showed subtle differences, this thesis concludes, in line with experts who emphasised the importance of parental voice rather than choice and in alignment to the values of democratic evaluation (MacDonald, 1974, 1976), schools' accountability to parents is not a one-way relation, but rather it is reciprocal. In other words, accountability could be seen as a shared responsibility. Therefore, envisaging the role of parents as 'outsiders' choosing schools for their children similar to how customers choose products sold in the marketplace leads to a limited understanding of parental role in the education system. The same goes for the role of the local community which, as MacDonald (1974, 1976) reminds us, can support both schools and more democratic approaches to school evaluation.

Last but not least, the current foundation of school evaluation, based on the *foundational ideas* (Christie and Alkin, 2013) of New Public Management and marketisation (Norris and Kushner, 2007), is at the heart of the problem with the current official approach, as it treats schools as linear organisations, ignoring their complexity, values, needs, and priorities. It is important to rethink the foundations of this approach in order to avoid the contradictions and 'perverse incentives' (Ned) embedded in the current system. The data which offered suggestions for rethinking the official approach has been guided by the second research question.

Research question 2 aimed to find out experts' views on alternatives to the official approach. Depending on the more pragmatic or more radical stance of the participants, the suggested alternatives ranged from minor changes to Ofsted's inspection methods to more radical alternatives which included reframing inspections, reconfiguring accountability, and reconceptualising school evaluation. The more radical alternatives, together with insights from evaluation theory, provided findings used in this thesis to recast the *pillars* of school evaluation.

Research question 3 aimed to ascertain how the alternatives suggested by experts could be assessed in the light of knowledge in the field of evaluation. This thesis concludes that the

polarisation of experts' views was aligned to Patton's (1996, 2006, 2011, 2012) distinction between traditional and developmental evaluation, as well as the underpinning view of school improvement as simple and linear or complex. This polarisation was also related to participants' orientation either to bureaucratic evaluation's focus on the values of the contractor (the Government) and autocratic evaluation's focus on 'objective' evidence, or orientation to democratic evaluation (MacDonald, 1974, 1976).

This thesis proposes a 'recasting' of the current *pillars* of school evaluation in England to include: voice, trusted professionals, capacity building, and collaboration. These alternative *pillars* rest on the *foundational ideas* (Christie and Alkin, 2013) of democratic values and community involvement, and the evaluation of schools as complex, non-linear organisations that operate in specific contexts. As detailed in Chapter 8, the main purpose of complexity-based evaluation should be collaborative learning because this is essential for improving both the single school and the whole education system.

## 9.3 Contribution to knowledge

This section explains how my research has sought to contribute to knowledge. First, in terms of Christie and Alkin's (2013) Evaluation Tree Theory, this thesis has sought to emphasise the importance of values in evaluation. Much evaluation either takes a value-free stance, as in autocratic evaluation (MacDonald, 1974, 1976) or is driven by contractual obligations (set by the Government). Whilst Christie and Alkin (2013) highlight the importance of 'valuing', by which they mean exercising judgement of quality (or value), values have been subsumed in their Evaluation Theory Tree under the branch of valuing. However, I would argue that this is a reductive perspective, and values would merit a separate branch, with a primary focus on questions related to values, such as: *Which values?* and *Whose values?* The work of UK-based theorists (MacDonald, 1974, 1976; Norris, 2015; Picciotto, 2015; Simons, 1987) can be seen as based on a *foundational idea* (Christie and Alkin, 2013) of democratic values (see Chapter 3, Section 3.4).

The contribution of this thesis specifically to the field of *school* evaluation rests on its unique conceptual framework, which draws concepts from several evaluation theories (developmental evaluation, democratic evaluation and Evaluation Theory Tree, see Chapter 3) to rethink the official English approach. The uniqueness of this conceptual framework is evidenced by advanced searches in ERIC. For example, a keyword search for 'Evaluation Theory Tree' and 'school' yielded no publications, whilst a search for 'democratic evaluation' and 'school'

yielded one journal article, reporting on research conducted in the US context (Howe et al., 2005). A keyword search for 'developmental evaluation' and 'school' yielded four journal articles reporting on US schools and one article reporting on developmental evaluation in schools in Northern Ireland (Balwin and Lander, 2019; Cooper et al., 2020; Peurach et al., 2016a; Peurach et al., 2016b; O'Connor, 2012). In the light of continued debates on school evaluation in England discussed in Chapter 2, employing insights from Patton (2006, 2011, 2012) and MacDonald (1974, 1976), combined with Christie and Alkin's (2013) focus on *foundational ideas*, could offer educational researchers helpful theoretical tools for rethinking school evaluation, in the post-pandemic world. Indeed, as reported by Ehren et al. (2020: 5), the pandemic led to a change in the purpose in school inspections in the direction of 'support and improvement' and 'liaison'. However, examining the *foundational ideas* of evaluation can reach even deeper, by questioning the very logic underlying approaches to school evaluation. Whilst the official approach in England has been driven, since the 1980s, by the *foundational ideas* of New Public Management, the pandemic has highlighted alternative educational priorities, including care, community and a common purpose instead of markets and competition.

Second, as noted in Chapter 1, the significance of this study also lies in the research design itself, in the form of expert (elite) interviews. As such, this study has also sought to provide some new insights into the experience of a female international junior researcher preparing for expert (elite) interviews and developing strategies to cope with the challenges of interviewing educational experts, both anticipated and unanticipated. It is hoped that the rich data collected in my study will encourage other junior researchers to conduct expert interviews, despite Dexter's (1964: 557) advice that elites are important people, and their time should not be taken up 'needlessly' by 'ill-prepared' novice researchers.

As explained in Chapter 3, the current definitions of elite (expert) interviews are relatively rare in educational research and, therefore, inconclusive about who might be 'experts' in the field of education. This study posits that educational experts are characterised by their in-depth historical, technical, theoretical, and/or practical knowledge and experience in the field of education rather than status and power. This is an important argument in the English context, as well as other systems where the political elites of government ministers and civil servants enjoy the status of education policymakers, whereas school practitioners are positioned as 'policy takers' (Gunter and Forrester, 2009). As can be seen from the data and findings reported

in Chapters 6-8, participants who at the time of data collection worked in the roles of *school practitioners* offered rich understandings and insights of equal value to those of university professors or officials working for Ofsted. In matters pertaining to detailed knowledge of individual children, the school and the local community, *school practitioners'* expertise could be, in fact, considered superior. This study recognises their voices and hopes that future research employs expert interviews to develop these ideas further.

As a junior researcher, I have also experienced a degree of unease at the stage of data analysis, upon noticing the strong polarisation of expert participants' views, reported in Chapters 6 and 7. This made me experience a lot of self-doubt and led to self-questioning: Was my analysis reductive? How could I make sense of this polarisation? Self-doubt continued until I reached two conclusions, firstly, that I need to trust my judgement and the data clearly pointed to a polarity of views. Secondly, that this polarisation could be a reflection of the 'real' world of politics, especially in two-party systems, where each 'side' has their own, distinctive, opposing point of view. The message to other junior researchers would be to trust their judgement and trust the data they have collected.

Third, the findings of this study have important implications for all key stakeholders in the system: *school practitioners*, *policy influencers* and *policy actors*, including education policymakers who hold government posts. One implication pertains to the importance of evaluation theory, as a source of methods employed in the evaluation of schools, the use of evaluation findings and the determination of the purpose of evaluation. As my study suggests, stakeholders and especially policymakers, need to carefully consider what they aim to achieve through evaluation. An evaluation system which targets the improvement of the individual school and employs competition as a driver of improvement, is not only counterproductive but may lead to an 'aggressive war' (Kelvin) between schools. Also, as Patton (2011) reminds us, making 'improvement' the ultimate purpose of evaluation is embedded in traditional evaluation, which may in turn lead to a number of negative consequences, such as a tendency to use reductive measures of improvement and, in the context of high-stakes tests and high-stakes accountability, game playing and the narrowing of educational provision. As explained by some expert participants (see Chapter 6), school inspection, conducted within a traditional approach to evaluation, cannot lead to improvement. Therefore, this study recommends Patton's (2006, 2011, 2012) developmental evaluation as a more beneficial approach, which can reinvigorate innovative practice and enhance the capacity of schools to engage in

meaningful self-evaluation and open dialogue with external evaluators.

Another implication pertains to two *pillars* of the official approach, namely standards and accountability. In relation to standards, the questions of *why* and *how* to use standards are important for all stakeholders to bear in mind, because standards can be used for good or ill (Chapter 8). In relation to accountability, how accountability is conceptualised and enacted matters a great deal as well. As discussed by Schedler (1999, see Chapter 3) and experts who advocated for a reconfiguration of accountability (Chapter 7), accountability can help or hinder schools from following their educational mission. By deconstructing the different rationalities and epistemologies underpinning the framings of school accountability, this thesis has sought to open up its conceptualisations and enactment to further debate.

Lastly, when key decision makers in the education system find themselves following a traditional approach to evaluation for over 30 years, as has been the case in England, it is time to honestly assess it and be open to adopting a new approach. This thesis offers a new approach, by drawing on developmental evaluation (Patton, 2006, 2011, 2012) and, in alignment with the radical stance of some of the experts who participated in this study, proposing alternative pillars of school evaluation based on 'trusted professionals', 'capacity building', 'voice', and collaboration. It is, therefore, hoped that this thesis contributes, in a modest way, to Waters and Brighouse's (2022: 23) argument for renewing 'hope, ambition and collaborative partnership'.

## 9.4 Limitations

This section reflects on the limitations of this study. All research studies have limitations, and this study is no exception. However, there is value to limitations as they may open doors for further research. First and foremost, whom I called 'expert' in this study could be challenged as limited or limiting. I took 'knowledge' to be at the core of expertise to draw a boundary around participants who could be viewed as experts, with their types of knowledge including historical, technical, theoretical, and/or practical knowledge gained through many years of working in different roles and at different levels of the education system. This was partially a subjective decision, which excluded from my research an important group of elite participants: policymakers working at the highest levels of decision making in the Department of Education and the Government. My decision might be challenged as limiting, especially if status and power were considered to be central to a definition of an 'expert' (Van Audenhove, 2017).

The sample of experts was purposive, benefiting from a snowballing technique (Cohen et al.,

2007). Although this sampling approach helped me to gain insight into diverse views, the findings of this study are limited to this sample of experts and cannot be generalised to a wider population of educational experts or education professionals. However, as discussed in Chapter 4, the purpose of this study was not to generate generalisable findings but to contribute to the field by offering some potentially generalisable insights. Because the knowledge about schools gleaned from different approaches to school evaluation is a social construct, I have sought a 'balanced' participant sample. As part of my purposive sampling (Cohen et al., 2007), I thus tried to identify experts who were supportive of the official approach to school evaluation, as well as experts who appeared to be critical of this approach. I did this by reviewing their work, if available (e.g., publications or public statements). Therefore, the findings from this study will hopefully be seen as presenting a more nuanced, complex picture of the diverse views and positions on the official approach to school evaluation in England and suggested alternatives. The ideas offered by experts who participated in this study could work as a possible resource informing a rethinking school evaluation in the English system.

As discussed in Chapter 3, I completed 15 expert interviews. The size of the sample may be seen as a limitation of this study. However, the number of interviews is not necessarily an indicator of the quality of research in expert interviews (Harvey, 2010) and my fieldwork generated rich, detailed responses from the participants, with the total wordage of the interview transcripts of 140,356 words and the total length of interviews of 21 hours. Gaining access to experts was one of the main challenges of my fieldwork. However, this limitation would be what one could expect in research involving expert interviews.

The inductive thematic data analysis occasionally made it challenging for me to handle the rich data I collected and to generate themes. However, this opened up opportunities in my interpretation of the data to look for other patterns. For instance, due to the polarisation of participants' views, it was difficult to find some common themes in the suggested alternatives to the current official approach. This made me revisit the data and findings in the light of evaluation theory (Christie and Alkin, 2013) and refocus the analysis of the data not only around the four *pillars* of the official approach but also the *foundational ideas* underlying this approach, as well as the ultimate purpose of evaluation within the official approach. Furthermore, there was an interview question which I decided to exclude from the analysis of data, pertaining to the methodologies underpinning school evaluation (see Appendix 2), because this question yielded relatively brief answers. This might be because, as one of my

participants noted, 'This is an interesting question. I've never thought about evaluation with quantitative and qualitative distinction' (Cooper). This might be also because, as another participant noted, 'it's very unhelpful to think about methodology in terms of qualitative and quantitative' (Orion) because 'we need both'. However, answers to the remaining interview questions presented in Chapters 6 and 7 have offered rich, insightful explanations by the participants.

## 9.5 Recommendations for further research

This section sets out suggestions for future research and provides some example research questions. The question about 'who is and who is not an educational expert' could merit further study, given the potential tension between power and status of some key educational decision-makers who have limited knowledge of education gained through professional experience or academic study. It would be also interesting to further explore the proposed recasting of the *pillars* of school evaluation. Also, researching the notions of 'development' and 'improvement' could extend Patton's work. Lastly, based on some participants' use of the terms of 'accountability' and 'inspection' as interchangeable, and on some experts' conflating evaluation with Ofsted inspection, it could be fruitful to study in more detail what educationalists mean by 'accountability' and 'evaluation' of schools.

## 9.6 Personal reflection on my research journey

As a researcher, hearing about the experiences and thoughts of experts has been a great privilege, as well as a challenge. I am acutely aware of the responsibility to accurately represent their positions and respect their confidentiality, due to the political controversies around the official approach to school evaluation. I have sought to enable their voices to be heard, as well as to carefully protect their identity. Through conducting this research from the social constructionist stance, my awareness of the complexity of school evaluation, including a whole range of unintended consequences of the official approach to school evaluation in England, has increased. This is well summarised by Box (1976) who points out that 'all models are wrong, but some are useful'. The claim that 'all models are wrong' implies a reductive, and therefore often misleading, nature of models. It is impossible to find a perfect model of school evaluation within a perfect approach. This experience will continue to shape my thinking and practice while I am contributing to the Turkish educational evaluation system. Whilst it is too early to provide details of my future practice, what has emerged for me from this study is an approach which is more attuned to the complexity of schools, which recognises the importance of

dialogue and trust between the evaluators and the evaluands and which acknowledges both the expected benefits and potential problems which may naturally emerge from the introduction of any evaluation and accountability mechanisms.

# REFERENCES

AEA-American Evaluation Association. (1994). *American Evaluation Association Guiding Principles for Evaluators*.

AEA-American Evaluation Association. (2018). *American Evaluation Association Guiding Principles for Evaluators.*

Alden, K. (2018). Creating Independent Learners: Placing Students at the Heart of the Assessment Process. *FORUM: For Promoting 3-19 Comprehensive Education*, 60(1), 45-50.

Alexander, R. J. (2012). Moral Panic, Miracle Cures and Educational Policy: What Can We Really Learn from International Comparison? *Scottish Educational Review*, 4-21.

Allen, R., and Vignoles, A. (2007). What Should an Index of School Segregation Measure? *Oxford Review of Education*, 33 (5), 643-668.

Altrichter, H., and Kemethofer, D. (2015). Does Accountability Pressure through School Inspections Promote School Improvement? *School Effectiveness and School Improvement*, 26(1), 32-56.

Andrews, T. (2012). What Is Social Constructionism? *Grounded Theory Review*, 11(1), 39-46.

Apple, M. W., and Beane, J. A. (1999). *Democratic Schools: Lessons from the Chalk Face.* Buckingham: Open University Press.

Baldwin, C.K., and Lander, R. (2019). Developmental Evaluator Functional Role Activities and Programmatic Developments: A Case Study Analysis. *American Journal of Evaluation*, 40(1), 35–54.

Ball, S. J., Maguire, M., and Braun, A. (2012). *How schools do policy: Policy enactments in secondary schools*. Routledge.

Ball, S.J. (2003). The Teacher's Soul and the Terrors of Performativity. *Journal of Education Policy*, 18(2), 215-228.

Barber, M. (2004). The Virtue of Accountability: System Redesign, Inspection, and Incentives in the Era of Informed Professionalism. *The Journal of Education*, 185(1), 7-38.

Barham, E. (1996). Central Control of the Curriculum and Teacher Involvement in Curriculum Change. *Journal of Curriculum and Supervision*, 12(1), 29-36.

Bates, A. (2016). Reculturing Schools in England: How 'Cult' Values in Education Policy Discourse Influence the Construction of Practitioner Identities and Work Orientations. *Critical Studies in Education*, 57 (2), 191-208.

Baxter, J. (2017). School Governor Regulation in England's Changing Education Landscape. *Educational Management Administration and Leadership*, 45(1), 20-39.

Beck, J. (2012). Reinstating knowledge: diagnoses and prescriptions for England's curriculum ills. *International Studies in Sociology of Education*, 22(1), 1-18.

Befani, B. (2020). *Choosing Appropriate Evaluation Methods*. CECAN.

BERA- British Educational Research Association (2018). *Ethical Guidelines for Educational Research, 4th Edition*. London. Retrieved From: Https://Www.Bera.Ac.Uk/Researchersresources/

Publications/Ethicalguidelines-For-Educational-Research-2018

Berliner, D. C. (2002). Educational Research: The Hardest Science of All. *Educational Researcher*, 31(8), 18-20.

Berry, J. (2009). Can there be an alternative to the centralized curriculum in England? *Improving Schools*, 12(1), 33-41.

Biesta, G. (2007). Why 'What Works' Won't Work: Evidence-Based Practice and the Democratic Deficit in Educational Research. *Educational Theory*, 57(1), 1-22.

Blok, H., Sleegers, P., and Karsten, S. (2008). Looking For a Balance Between Internal and External Evaluation of School Quality: Evaluation of the SVI Model. *Journal of Education Policy*, 23(4), 379-395.

Bogner, A., Littig, B., and Menz, W. (2018). Generating Qualitative Data with Experts and Elites. *The SAGE Handbook of Qualitative Data Collection*, 652-667.

Box, G. (1976). Science and Statistics. *J American Stat Assoc* 71(356), 791-799.

Box, R. C., Marshall, G. S., Reed, B. J., and Reed, C. M. (2001). New Public Management and Substantive Democracy. *Public Administration Review*, 61(5), 608-619.

Brady, A. M. (2016). The Regime of Self-Evaluation: Self-Conception for Teachers and Schools. *British Journal of Educational Studies,* 64(4), 523-541.

Braun, V., and Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, 3(2), 77-101.

Brighouse, T., and Waters, M., 2022). *About Our Schools: Improving on Previous Best*. Crown House Publishing Ltd.

Brimblecombe, N., Ormston, M., and Shaw, M. (1995). Teachers' Perceptions of School Inspection: A Stressful Experience. *Cambridge Journal of Education*, 25(1), 53-61.

Brundrett, M. (2015). Policy on the Primary Curriculum Since 2010: The Demise of the Expert View. *London Review of Education*, 13(2), 49-59.

Brundrett, M., and Rhodes, C. (2011). *Leadership For Quality and Accountability in Education*. London: Routledge.

Bubb, S., and Earley, P. (2008). *From Self-Evaluation to School Improvement: The Importance of Effective Staff Development*. Reading: Cfbt Education Trust.

Burgess S., Wilson D, Lupton R. (2005). Parallel Lives? Ethnic Segregation in Schools and Neighbourhoods. *Urban Studies*, 42(7), 1027-1056.

Burgess, S., Greaves, E., Vignoles, A., and Wilson, D. (2015). What Parents Want: School Preferences and School Choice. *The Economic Journal*, 125(587), 1262-1289.

Burgess, S., Mcconnell, B., Propper, C., and Wilson, D. (2007). The Impact of School Choice on Sorting by Ability and Socioeconomic Factors in English Secondary Education. *Schools and the Equal Opportunity Problem,* 273.

Burr, V. (2015). *Social Constructionism*. London and New York: Routledge.

Campbell, D. (1957). Factors Relevant to the Validity of Experiments in Social Settings. *Psychological Bulletin*, 54, 297-312.

Carden, F., and Alkin, M. C. (2012). Evaluation Roots: An International Perspective. *Journal of Multidisciplinary Evaluation*, 8(17), 102-118.

Challenge Partners (2015). Annual Report 2014. London: Challenge Partners [Online]. Available At: Http://Challengepartners.Org/Uploads/Resources_One/1426520332-CP_-_Annual_ Report_-_2015_-_FINAL.Pdf [Accessed: 1 July 2016].

Chamberlain, K., and Hodgetts, D. (2018). Collecting Qualitative Data with Hard-to-Reach Groups. *Handbook of Qualitative Data Collection*. London: SAGE Publications Ltd.

Chapman, C. (2000). Improvement, Inspection and Self-Review. *Improving Schools*, 3(2), 57-63.

Chapman, C., and Earley, P. (2010). School Inspection/External School Evaluation. In P. Peterson, E. Baker, and B. Mcgaw (Eds.), *International Encyclopedia of Education*. Oxford: Elsevier Science.

Chapman, C., and Sammons, P. (2013). School Self-Evaluation for School Improvement: What Works and Why? *Cfbt Education Trust*. England.

Chitty, C. (2014). *Education Policy in Britain*. Palgrave Macmillan.

Christie, C. A., and Alkin, M. C. (2013). An evaluation theory tree. *Evaluation roots: A wider perspective of theorists' views and influences,* 2, 11-57.

Chychuk, A. (2015). The system of teacher education management in Great Britain. *Порівняльна професійна педагогіка,* 5 (4), 80-86.

Cohen, L., Manion, L., and Morrison, K. (2007). *Research Methods in Education*. Routledge Taylor and Francis Group: London and New York.

Courtney, S. J. (2012). Ofsted's Revised School Inspection Framework: Experiences and Implications. BERA Conference, Manchester.

Cousins, J. B., and Ryan, K. E. (2009). Dilemmas For Educational Evaluation in A Globalized Society. In K. Ryan, and J.B. Cousins (Eds.), *The Sage International Handbook of Educational Evaluation*. Thousand Oaks, CA: Sage.

Daugherty, R., Black, P., Ecclestone, K., James, M., and Newton, P. (2008). Alternative perspectives on learning outcomes: Challenges for assessment. *Curriculum journal*, 19(4), 243-254.

Davis, H. and Martin, S. (2008). The Future of Public Services Inspection. In H. Davis, and S. Martin (Eds.), *Public Services in the UK*. London: Jessica Kingsley.

Davis, J. (2002). The Inner London Education Authority and The William Tyndale Junior School Affair, 1974-1976. *Oxford Review of Education* 28.2/3, *A Century of Local Education Authorities*, June-September, 275-98.

Day, C., and Sammons, P. (2013). *Successful Leadership: A Review of the International Literature*. Cfbt Education Trust. England

Dexter, L. A. (2006). *Elite and Specialized Interviewing*. ECPR Press.

DfE- Department for Education (2010). *The Importance of Teaching*. London: The Stationery Office.

DfES- The Department for Education and Skills (1980). *A Framework for the School Curriculum.*

DfES- The Department for Education and Skills (2004). A New Relationship with Schools. London: DfES.

Dunford, J. E. (2017). *Her Majesty's Inspectorate of Schools Since 1944: Standard Bearers or Turbulent Priests.* Routledge.

Dupuis, D. R., Baxter, C., and Dobson, J. (2013). Book Review: Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use Patton, MQ (2011). New York: Guilford Press, vii-375. *Journal of Community and Applied Social Psychology*, 23, 258-260.

Earley, P. (2013). *Exploring the School Leadership Landscape: Changing Demands, Changing Realities*. Bloomsbury Publishing.

Ehren, M. C. (2016). *Methods and Modalities of Effective School* Inspections. Dordrecht: Springer International Publishing.

Ehren, M. C. M. and Godfrey, D. (2017). External Accountability of Collaborative Arrangements: A Case Study of a Multi Academy Trust in England. *Educational Assessment, Evaluation and Accountability*, 29(4), 339-362.

Ehren, M. C., Jones, K., and Perryman, J. (2016). Side Effects of School Inspection; Motivations and Contexts for Strategic Responses. In M. C. M. Ehren (Ed.), *Methods and Modalities of Effective School Inspections. Accountability and Educational Improvement*. Cham, Germany: Springer.

Ehren, M., Chapman, C., and Montecinos, C. (2020). *COVID-19: do we need to reimagine the purpose of school inspections.*

Elliott, A. (2012). Twenty Years Inspecting English Schools: Ofsted 1992-2012.

Elliott, J. (1993). What Have We Learned from Action Research in School-Based Evaluation? *Educational Action Research*, 1(1), 175-186.

Elliott, J. (2001). Making Evidence-Based Practice Educational. *British Educational Research Journal*, 27 (5), 555-574.

Elliott, J. (2007). A Reply to Hargreaves. In M. Hammersley (Ed.), *Educational Research and Evidence-Based Practice.* Sage.

Elmore, R. (2006). *Leadership As the Practice of Improvement. Paper Prepared for the International Conference on Perspectives on Leadership for Systemic Improvement*. Sponsored by the Organisation for Economic Cooperation and Development (OECD).

Elwood, S. A., and Martin, D. G. (2000). "Placing" Interviews: Location and Scales of Power in Qualitative Research. *The Professional Geographer*, 52(4), 649-657.

Exley, S., and Ball, S. J. (2011). Something Old, Something New... Understanding Conservative Education Policy. *The Conservative Party and Social Policy*, 97-118.

Fagen, M. C., Redman, S. D., Stacks, J., Barrett, V., Thullen, B., Altenor, S., and Neiger, B. L. (2011). Developmental Evaluation: Building Innovations in Complex Environments. *Health Promotion Practice*, 12(5), 645-650.

Fidler, B., Russell, S., and Simkins, T. (1997). *Choices For Self-Managing Schools: Autonomy and Accountability*. London, Paul Chapman Publishing.

Fisher, T. (2008). The Era of Centralisation: The 1988 Education Reform Act and its Consequences. *FORUM: For Promoting 3-19 Comprehensive Education*, 50(2), 255-261.

Fitz-Gibbon, C., and Stephenson, N. J. (1996). *Inspecting Her Majesty's Inspectors: Should Social Science and Social Policy Cohere*. In European Conference on Educational Research, Seville, Spain.

Fullan, M. (1991). *The New Meaning of Educational Change*. London/ New York: Cassell.

Gergen, K.J. (2009). *An Invitation to Social Construction*. London: Sage Publications Ltd.

Gewirtz, S. Bowe R., and Ball S. (1994). *Choice, Competition and Equity: Lessons from Research in the UK*.

Gilbert, C. (2012). *Towards A Self-Improving System: The Role of School Accountability*. Nottingham: National College of School Leadership.

Gillard, D. (2011). Education in England: A Brief History.

Glasswell, K., and Ryan, J. (2017). Reflective Practice in Teacher Professional Standards: Reflection as Mandatory Practice. In R. Brandenburg, K. Glasswell, M. Jones, and J. Ryan, et al (Eds.), *Reflective Theory and Practice in Teacher Education*. Springer, Singapore.

Goldacre, B. (2013). *Building Evidence into Education*.

gov.uk (1984). *Parental Influence at School: A New Framework for Schools in England and Wales*. Green Paper.

gov.uk (2020). *Magenta Book: Central Government Guidance on Evaluation*. HM Treasury.

Gray, S. L. (2007). Teacher as Technician: Semi-Professionalism after the 1988 Education Reform Act and Its Effect on Conceptions of Pupil Identity. *Policy Futures in Education*, 5(2), 194-203.

Greany, T., and Higham, R. (2018). *Hierarchy, Markets and Networks: Analysing the Self-Improving School-Led System' Agenda in England and the Implications for Schools.*

Greene, J. (2006). Evaluation, Democracy, and Social Change. In I. Shaw, I.G.R. Shaw, J.C. Greene and M.M. Mark (Eds.), *the SAGE Handbook of Evaluation*. SAGE Publications.

Greene, J. C. (2009). Meaningfully Engaging with Difference Through Mixed Methods Educational Evaluation. In K. Ryan and J.B. Cousins (Eds.), *The Sage International Handbook of Educational Evaluation*.

Guijt, I. M., Kusters, C. S. L., Lont, H., and Visser, I. (2012). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*, 12-015. Wageningen UR Centre For Development Innovation.

Gunter, H. M., and Forrester, G. (2009). School Leadership and Education Policymaking in England. *Policy Studies,* 30(5), 495-511.

Hacking, I. (1999). *The Social Construction of What?* Harvard University Press.

Hall, C., and Noyes, A. (2007). *The Impact of School Self-Evaluation Processes on British Teachers. Views of Their Work and Professionalism.* Paper Presented at the BERA, London.

Hammersley, M. (2007). Educational Research and Teaching: A Response to David Hargreaves' TTA Lecture. In M. Hammersley (Ed.), *Educational Research and Evidence-Based Practice*. Sage.

Hammersley-Fletcher, L. (2007). Morality, Power and Leadership--Primary School Perspectives. *Management in Education*, 21(4), 19-24.

Hargreaves, D. H. (2007). Teaching As a Research-Based Profession: Possibilities and Prospects. In M. Hammersley (Ed.), *Educational Research and Evidence-Based Practice*. Sage.

Hargreaves, D. H. (2010). *Creating A Self-Improving School System*.

Harris, A., and Chapman, C. (2004). Improving Schools in Difficult Contexts: Towards A Differentiated Approach. *British Journal of Educational Studies*, 52(4), 417-431.

Harvey, W. S. (2010). Methodological Approaches for Interviewing Elites. *Geography Compass* 4(3): 193-205.

Harvey, W. S. (2011). Strategies For Conducting Elite Interviews. *Qualitative Research*, 11(4), 431-441.

Hatcher, R. (2011). The Conservative-Liberal Democrat Coalition Government's "Free Schools" In England. *Educational Review*, 63(4), 485-503.

Hellawell, D. E. (1992). Structural Changes in Education in England. *International Journal of Educational Reform*, 1(4), 356-365.

Hicks, T. (2015). Inequality, Marketisation and the Left: Schools Policy in England and Sweden. *European Journal of Political Research,* 54(2), 326-342.

HMI- Her Majesty's Inspectorate (1970). *HMI Today and Tomorrow*. Http://Www.Educationengland.Org.Uk/Documents/Hmi/1970-Today-and-Tomorrow.Html

HMI- Her Majesty's Inspectorate (2016/7). *The Annual Report of Her Majesty's Chief Inspector of Education*, Children's Services and Skills 2016/17.

Hofer, S. I., Holzberger, D., and Reiss, K. (2020). Evaluating School Inspection Effectiveness: A Systematic Research Synthesis On 30 Years of International Research. *Studies In Educational Evaluation*, 65, 100864.

Hogan, R.L., 2007. The Historical Development of Program Evaluation: Exploring Past and Present. *Online Journal for Workforce Education and Development*, *2*(4), 5.

Holstein, J. A., and Gubrium, J. F. (2008). *Handbook of Constructionist Research*. Guilford Publications.

Hopkins, D., Ainscow, M., and West, M. (1994). *School Improvement in An Era of Change*. London: Cassell.

Hosking, D. M. (2008). Can Constructionism Be Critical. In J. A. Holstein, and J. F. Gubrium (Eds.), *Handbook of Constructionist Research*. Guilford Publications.

Hoskins, K. (2012). Raising Standards 1988 To the Present: A New Performance Policy Era *Journal of Educational Administration and History*, 44(1), 5-19.

House, E., and Howe, K. R. (1999). *Values in Evaluation and Social Research*. Thousand Oaks, CA: Sage.

Howe, K. R. (2003). *Closing Methodological Divides: Toward Democratic Educational Research*. Philosophy and Education, 11. Boston, MA: Kluwer.

Howe, K.R., and Ashcraft, C. (2005). Deliberative Democratic Evaluation: Successes and Limitations of an Evaluation of School Choice. *Teachers College Record*, 107(10), 2275–2298.

Hudson, C. (2007). Governing the governance of education: The state strikes back? *European Educational Research Journal*, *6*(3), 266-282.

Janssens, F. J. G, and Van Amelsvoort, G. H. W.C.H. (2008). School Self-Evaluations and School Inspections in Europe: An Exploratory Study. *Studies In Educational Evaluation*, 34(1), 15-23.

Janssens, F. J., and Ehren, M. C. (2016). Toward A Model of School Inspections in A Polycentric System. Evaluation and Program Planning, 56, 88-98.

Joint Committee on Standards for Educational Evaluation. (1988). The Personnel Evaluation Standards. Newbury Park, CA: Sage.

Joint Committee on Standards for Educational Evaluation. (1994). The Program Evaluation Standards. Thousand Oaks, CA: Sage.

Jones, K., and Tymms, P. (2014). Ofsted's Role in Promoting School Improvement: The Mechanisms of the School Inspection System in England. *Oxford Review of Education*, 40(3), 315-330.

Kushner, S., and Norris N. (2007). *Dilemmas of Engagement: Evaluation and the New Public Management*. Elsevier.

Lane, J. E. (2000). *New Public Management*. London: Routledge.

Learmonth, J. (2020)*. Inspection: What's in It for Schools*? Routledge.

Lee, J., and Fitz, J. (1997). HMI and Ofsted: Evolution or Revolution in School Inspection? *British Journal of Educational Studies,* 45 (1).

Levin-Rozalis, M., Rosenstein, B., and Cousins, B. (2009). Chapter 11 | A Precarious Balance: Educational Evaluation Capacity Building in A Globalized Society. *In K*. Ryan, and J.B. Cousins (Eds.), *The Sage International Handbook of Educational Evaluation*. Thousand Oaks, CA: Sage.

Levitt, R., Martin, S. J., Nutley, S., and Solesbury, W. (2010). *Evidence For Accountability: Using Evidence in the Audit, Inspection and Scrutiny of UK Government*. London: Nuffield Foundation.

Lindblom, C. E. (1979). Still Muddling, Not Yet Through. *Public Administration Review*, 39(6), 517-526.

Loseke, D. R., and Kusenbach, M. (2008). The Social Construction of Emotion. In J. A. Holstein and J. F. Gubrium (Eds.), *Handbook of Constructionist Research.* Guilford Publications.

Macbeath, J. (2005). *School Self-Evaluation Background, Principles, and   Key Learning*. National College for School Leadership.

Macbeath, J. (2006). *School Inspection and Self-Evaluation: Working with the New Relationship.* Routledge.

MacDonald, A., Reilly, J., and Worsley, L. (2016). There Is Another Way: Building a New Vision for Schools from the Bottom Up. *FORUM: for promoting 3-19 comprehensive education*, 58(2), 195-204.

Macdonald, B. (1974). Evaluation and the Control of Education. In Macdonald, B. and Walker, R. (Eds.). *Safari I: Innovation, Evaluation, Research and the Problem of Control*. Norwich, Centre for Applied Research in Education, University of East Anglia, 9-22. Also In Tawney, D. (Ed.). (1976).

Curriculum Evaluation Today: Trends and Implications, Schools Council Research Studies, London, Macmillan Educational.

Macdonald, B. (1979). *Democracy and Evaluation*. CMI Occasional Paper Series, 1.

Matthews, P., and Ehren, M.C.M. (2017). Accountability and Improvement in Self-Improving School Systems. In P. Earley and T. Greany (Eds.), *School Leadership and Education System Reform*. Bloomsbury Publishing.

Matthews, P., and Headon, M. (2015). *Multiple Gains: An Independent Evaluation of Challenge Partners. Peer Reviews of Schools*. London: UCL-Institute of Education, Trentham Press.

Matthews, P., and Sammons, P. (2004). *Improvement Through Inspection*. 2004, July. London: Ofsted.

Matthews, P., and Smith, G. (1995). OFSTED: Inspecting Schools and Improvement Through Inspection. *Cambridge Journal of Education*, 25 (1), 23-34.

Mccluskey, G. (2017). Mapping, Measuring and Monitoring Achievement: Can A New Evaluation Framework Help Schools Challenge Inequalities? *Improving Schools,* 20(1), 5-17.

Mcdowell, L. (1998). Elites In the City of London: Some Methodological Considerations. *Environment and Planning A,* 30, 2133-2146.

Mcnamara, G., and O'Hara, J. (2009). Where Global Meets Local: Contexts, Constraints and Consensus in School Evaluation in Ireland. *Sage International Handbook on Educational Evaluation*, 273-291. Thousand Oaks: Sage.

Mirick, R.G. and Wladkowski, S.P., (2019). Skype in qualitative interviews: Participant and researcher perspectives. *The Qualitative Report,* 24(12), 3061-3072.

Namey, E., Guest, G., Thairu, L., and Johnson, L. (2008). Data Reduction Techniques for Large Qualitative Data Sets. In Guest G., and Macqueen K.M. *Handbook for Team-Based Qualitative Research*, 2(1), 137-1. Altamira Press.

Nelson, R., and Ehren, M. (2014). *Review and Synthesis of Evidence on the (Mechanisms of) Impact of School Inspections.* Utrecht, the Netherlands: Dutch Inspectorate of Education.

Nelson, R., Ehren, M.C.M., and Godfrey, D. (2015). *Literature Review on Internal Evaluation*. London: UCL Institute of Education.

Nevo, D. (2009). Accountability and Capacity Building: Can They Live Together. In K.E. Ryan and J.B. Cousins (Eds.), *The SAGE International Handbook of Educational Evaluation*. Sage Publications.

Nfer- National Foundation for Educational Research (2009). *Teacher Resignation and Recruitment Survey*. London: NFER.

Norris, N. (2007). Evaluation and Trust. In S. Kushner, and N. Norris (Eds.), *Dilemmas of Engagement: Evaluation and the New Public Management*. Elsevier.

Norris, N., and Kushner, S. (2007). The New Public Management and Evaluation. In Kushner, and N. Norris (Eds.), *Dilemmas of Engagement: Evaluation and the New Public Management*. Elsevier.

Norris, N. (2015). Democratic Evaluation: The Work and Ideas of Barry Macdonald. *Evaluation*, 21(2), 135-142.

O'Connor, U. (2012). Schools together: Enhancing the citizenship curriculum through a non-formal

education programme. *Journal of Peace Education,* 9(1), 31-48.

OECD- Organisation for Economic Co-Operation and Development (2015). *Education Policy Outlook 2015: Making Reforms Happen.* OECD Publishing.

Ofsted (1998). *School Self-Evaluation Matters.* London, Ofsted

Ofsted (2003). *Framework For Inspecting Schools.* London, Ofsted

Ofsted (2005). *Primary Self-Evaluation Form.* London: Dfes.

Ofsted (2015). *The Common Inspection Framework: Education, Skills and Early Years.*

Ofsted (2018). *Annual Teachers Survey 2018: Teachers' Awareness and Perceptions of Ofsted.*

Ofsted (2019a). *The Education Inspection Framework.*

Ofsted (2019b). Teachers' Awareness and Perceptions of Ofsted.

Owens, L. L. (2022). An Implausible Virtual Interview: Conversations with a Professional Research Subject. *Sociological Methodology*, 52(2), 121-140.

Ozga, J. (1995). Deskilling A Profession: Professionalism, Deprofessionalisation and the New Managerialism. In H. Busher and R. Saran (Eds.), *Managing Teachers as Professionals in Schools.* Routledge.

Ozga, J., and Lawn, M. (2014). Inspectorates and Politics: The Trajectories of School Inspection in England and Scotland. *Revue Française De Pédagogie. Recherches En Éducation*, 186, 11-21.

Ozsezer, S. (2016). A *Case Study on Peer Review in Schools: The Perspectives of Senior Leaders Who Participated in the Challenge Partner's Quality Assurance Peer Review* (Master Dissertation). University College London Institute of Education, London.

Pagett, L. (1996). Assessment: A Contemporary Perspective from An English Primary School. *Journal of Curriculum and Supervision*, 12(1), 21-28.

Parliament, U. K. (1988). *Education Reform Act 1988.* London: HM Stationary Office.

Parliament, U. K. (2005). *The Education Act 2005.* London: HM Stationary Office.

Patton, M. Q. (1994). Developmental Evaluation. *Evaluation Practice*, 15(3), 311-319.

Patton, M. Q. (1996). A World Larger Than Formative and Summative. *Evaluation Practice*, 17(2), 131-144.

Patton, M. Q. (2002). Two Decades of Developments in Qualitative Inquiry: A Personal, Experiential Perspective. *Qualitative Social Work*, 1(3), 261-283.

Patton, M. Q. (2006). Evaluation For the Way We Work. *Nonprofit Quarterly*, 13(1), 28-33.

Patton, M. Q. (2011). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use.* Guilford Press.

Patton, M. Q. (2012). Developmental Evaluation for Equity-Focused Evaluations. In M. Bamberger and M. Segone (Eds.), *Evaluation for Equitable Development Results.* New York, UNICEF.

Patton, M. Q. (2021). Emergent Developmental Evaluation Developments. *Journal of Multidisciplinary Evaluation*, 17(41), 23-34.

Pawson, R. (2013). *The Science of Evaluation: A Realist Manifesto*. Sage.

Pawson, R., and Tilley, N. (1997). *Realistic Evaluation*. Sage.

Peer Challenge (nd). The Medway Teaching School Alliance - Peer Review and Challenge. Guidance for Governing Bodies.

Penninckx, M., and Vanhoof, J. (2017). What Stated Aims Should School Inspection Pursue? - Views of Inspectors, Policy-Makers, and Practitioners. In J. Baxter (Ed.), *School Inspectors: Policy Implementers, Policy Shapers in National Policy Contexts*. Springer.

Perryman, J. (2006). Panoptic Performativity and School Inspection Regimes: Disciplinary Mechanisms and Life Under Special Measures. *Journal of Education Policy*, 21(2), 147-161.

Perryman, J. (2009). Inspection and the Fabrication of Professional and Performative Processes. *Journal of Education Policy*, 24(5), 611-631.

Peurach, D.J., Glazer, J.L. and Winchell, Lenhoff, S. (2016). The Developmental Evaluation of School Improvement Networks. *Educational Policy,* 30(4), 606–648.

Peurach, D.J., Lenhoff, S.W. and Glazer, J.L. (2016). Large-Scale High School Reform through School Improvement Networks: Exploring Possibilities for "Developmental Evaluation". *Teachers College Record*, 118(13).

Pfohl, S. (2008). The Reality of Social Constructions, 669-686. In J.A. Holstein and J.F. Gubrium (Eds.), *Handbook of Constructionist Research*. Guilford Publications.

Picciotto, R. (2015). Democratic Evaluation for the 21st Century. *Evaluation*, *21*(2), 150-166.

Plowden, B. B. H. P. (1967). *Children and their Primary Schools: A Report of the Central Advisory Council for Education (England)*, 1. HM Stationery Office.

Power, M. (1997). *The Audit Society: Rituals of Verification*. OUP Oxford.

Purdue, D. (2005). Performance Management for Community Empowerment Networks. *Public Money and Management.* 25 (2), 123-130.

PwC- Price Water House Coopers (2007). *Independent study into school leadership*. London, UK: Department for Education and Skills.

Ranson, S. (2007). Public accountability in the age of neo-liberal governance. In B. Lingard. and J. Ozga (Eds.), *The RoutledgeFalmer reader in education policy and politics*. Routledge.

Richards, C. (2015). More Outstanding Nonsense: A Critique of Ofsted Criteria. *FORUM: For Promoting 3-19 Comprehensive Education,* 57(2), 233-238.

Richards, D. (1996). Elite Interviewing: Approaches and Pitfalls. *Politics,* 16(3): 199-204.

Richardson, M. (2022). *Rebuilding Public Confidence in Educational Assessment*. UCL Press.

Richmond, J. (1996). Quantitative Measures of Secondary School Performance Using School-Level Data. *Educational Management and Administration*, 24(2), 151-162.

Richmond, J. (2017). Assessment in English 3 to 11. *FORUM: For Promoting 3-19 Comprehensive Education,* 59(2), 133-140.

Ritchie, R. (2007). School Self-Evaluation. In S. Kushner, and N. Norris (Eds.), *Dilemmas of Engagement: Evaluation and the New Public Management*. Elsevier.

Roberts, N., and Hill, R. (2019). *School Inspections in England: Ofsted*. House of Commons Library, Briefing Paper, 07091.

Rosenthal, L. (2004). Do School Inspections Improve School Quality? Ofsted Inspections and School Examination Results in the UK. *Economics of Education Review,* 23 (2), 143-151.

Ryan, K., and Cousins J.B. (2009). Part III | Educational Evaluation, Capacity Building, and Monitoring. In K. Ryan and J.B. Cousins (Eds.), *The Sage International Handbook of Educational Evaluation*. Sage Publications.

Ryan, K., and Feller, I. (2009). Evaluation, Accountability and Performance: Measurement in National Education Systems. In K. Ryan and J.B. Cousins (Eds.), *The Sage International Handbook of Educational Evaluation*. Sage Publications.

Schedler, A. (1999). *Conceptualizing Accountability. The Self-Restraining State: Power and Accountability in New Democracies.*

Schleicher, A. (2014). *Equity, Excellence, and Inclusiveness in Education: Policy Lessons from Around the World.* OECD Publishing.

Scriven, M. (1967). The Methodology of Evaluation. In R. W. Tyler, R. M. Gagné, and M. Scriven (Eds.), *Perspectives of Curriculum Evaluation*, 39-83. Chicago, IL: Rand Mcnally.

Scriven, M. (1997). Empowerment Evaluation Examined. *Evaluation Practice*, 18(2), 165-175.

Shadish, W.R. and Luellen, J.K. (2005). History of Evaluation. In Mathison S. *The Encyclopaedia of Evaluation*. Thousand Oaks, CA: Sage.

Shaw, I., Newton, D., Aitkin, M. and Darnell, R. (2003). Do Ofsted Inspections Make a Difference to GCSE Results? *British Journal of Educational Research,* 29 (1).

Shaw, S. (2011). History of Education. In H. Cooper and Elton-Chalcraft (Eds.), *Professional Studies in Primary Education.* Sage

Simons, H. (1987). *Getting To Know Schools in A Democracy: The Politics and Process of Evaluation.* Routledgefalmer: London and New York.

Simons, H. (2006). Ethics and Evaluation. In I.F. Shaw, J.C. Greene and M.M. Mark (Eds,), *The International Handbook of Evaluation*. London: Sage.

Simons, H. (2009). Case Study Research in Practice. In Simons, H. (Ed.), *Case Study Research in Practice.* SAGE Publications.

Sriven, M. (1965). *The Methodology of Evaluation*. Bloomington: Indiana University.

Stake, R. (1967). The Countenance of Educational Evaluation. *Teachers College Record*, 68, 523-540.

Stake, R. E. (1986). *Quieting Reform: Social Science and Social Action in An Urban Youth Program.* Urbana and Chicago: University of Illinois Press.

Stake, R. E. (2005). Qualitative Case Studies, In N.K. Denzin, and I.S. Lincoln (Eds.), *The Sage Handbook of Qualitative Research*. Thousand Oaks: Sage Publications.

Strain, M. (2009). The Education Reform Act 1988: Success or Failure? A Short Report of A BELMAS Discussion Forum. *Management in Education*, 23(4), 151-154.

Sturges, J. E., and Hanrahan, K. J. (2004). Comparing telephone and face-to-face qualitative interviewing: a research note. *Qualitative research*, 4(1), 107-118.

Suchman, E. (1967). *Evaluative Research: Principles and Practice in Public Service and Social Action Programs*. New York, NY: Russell Sage.

Tan, C. Y. A., Swe, K. M. M., and Poulsaeman, V. (2021). Online examination: a feasible alternative during COVID-19 lockdown. *Quality Assurance in Education*.

Thomas, G. (1999). Standards and School Inspection: The Rhetoric and the Reality. In C. Cullingford (Ed.), *An inspector calls-OFSTED and its effect on school standards*. Kogan.

Thomas, G. (2013). *How To Do Your Research Project*. London: Sage Publications.

Thrupp, M. (2005). *School Improvement: An Unofficial Approach*. Continuum: London, New York.

Torrance, H. (2011). Using Assessment to Drive the Reform of Schooling: Time to Stop Pursuing the Chimera? *British Journal of Educational Studies*, 59(4), 459-485.

Townsend, T. and Avalos, B. (2007). *International Handbook of School Effectiveness and Improvement*. Dordrecht: Springer.

Trafford, V. and Leshem, S. (2008). *Stepping stones to achieving your doctorate: By focusing on your viva from the start*. McGraw-Hill Education (UK)

UEA- University of East Anglia (2013). *Research Ethics Policy.* Retrieved From: Https://Www.Uea.Ac.Uk/Documents/6207125/8216919/University+Research+Ethics+Polic

UKES- United Kingdom Evaluation Society (2013a). *UKES Guidelines for Good Practice in Evaluation*.

UKES- United Kingdom Evaluation Society (2013b). *UKES Evaluation Capabilities Framework.*

Van Audenhove L. (2017). *Expert and Elite Interviews in Social Science*. Available At: Https://Www.Researchgate.Net/Publication/317185080_Expert_and_Elite_Interviews_In_The_SOcia l_Sciences [23.04.2018]

Volante, L. (2005). Accountability, Student Assessment, and the Need for A Comprehensive Approach. *International Electronic Journal for Leadership in Learning*, 9(6).

Walford, G. (1994). Reflections On Researching the Powerful. Researching the Powerful in Education, 222-231.

Welch, C., Marschan-Piekkari, R., Penttinen, H., and Tahvanainen, M. (2002). Corporate Elites as Informants in Qualitative International Business Research. *International Business Review*, 11(5), 611-628.

West, A., and Wolfe, D. (2018). *Academies, the School System in England and a Vision for the Future.*

Wikeley, F. and Hughes, M. (1995). Parents and Educational Reform in the UK. *International Journal*

*of Educational Reform*, 4(3), 301-309.

Wilcox, B. and Gray, J. (1996). *Inspecting Schools*. Buckingham, Open University Press.

Wilkins, A. (2016). *Modernising school governance: Corporate planning and expert handling in state education*. Routledge.

Wilkins, A. (2016). *Modernising School Governance: Corporate Planning and Expert Handling in State Education.* Routledge.

Wilkins, A. (2017). Rescaling the Local: Multi-Academy Trusts, Private Monopoly and Statecraft in England. *Journal of Educational Administration and History*, 49(2), 171-185.

Zuckerman, H. A. (1972). Interviewing An Ultra-Elite. *The Public Opinion Quarterly*, 36, 159- 175.

# APPENDICES

## Appendix 1 Ethical approval

**Appendix 1a Participant Information Statement and consent form for participants who are unlikely to be identifiable**



**Sevda** OZSEZER
Research Student
School of Education and Lifelong
Learning

University of East Anglia

**Faculty of Social Science**

University of East Anglia Norwich Research Park Norwich NR4 7TJ United Kingdom

Email: S.Ozsezer@uea.ac.uk Tel: +44 (0)1603 591451

*Approaches to school evaluation in England since 1988*

*PARTICIPANT INFORMATION STATEMENT*

### 1. What is this study about?

I would like to invite you to take part in my PhD research project about approaches to school evaluation in England since 1988. My research focuses on the changes to school evaluation since the 1988 Education Reform Act (ERA) and their underpinning methodologies to consider whether 'lessons from England' might be relevant to other education systems including the Turkish education system. You are being invited to take part in this research because you are an expert in the field.

This Participant Information Statement tells you about the research study. Knowing what is involved will help you decide if you want to take part in the study. Please read this sheet carefully and ask questions about anything that you don't understand or want to know more about.

Participation in this research study is voluntary. By giving consent to take part in this study you are telling us that you:
Understand what you have read.
        YES/NO Agree to take part in the research study as outlined below.
        YES/NO Agree to the use of your personal information as described.

### 2. Who is running the study?
Sevda ÖZSEZER, a PhD student in the School of Education and Lifelong Learning at the University of East Anglia (UEA), is running the study under the supervision of Dr Agnieszka Bates and Professor Nigel Norris.

### 3. What will the study involve for me?

I would like to interview you and, if possible, to audio-record our interview. The interview will seek your views on the changing approaches to school evaluation in England. The interview will be arranged at a time and place convenient for you. I will be happy to meet either in your workplace or a public place such as a café.

**4. How much of my time will the study take?**
The interview will last approximately an hour.

**5. Do I have to be in the study? Can I withdraw from the study once I've started?**
Being in this study is completely voluntary and you do not have to take part. Your decision whether to participate will not affect your current or future relationship with the researchers or anyone else at the University of East Anglia. If you decide to take part in the study and change your mind later, you are free to withdraw at any time up until I have collected the data and analysed the results. You are also free to stop the interview at any time. Should you wish to withdraw from the interview, any recordings will be erased and the information you have provided will not be included in the study results. You may also refuse to answer any questions that you do not wish to answer during the interview.

**6. Are there any risks or costs associated with being in the study?**
Aside from giving up your time, we do not expect that there will be any risks or costs associated with taking part in this study.

**7. Are there any benefits associated with being in the study?**
This study will seek to draw on your knowledge and expertise to enhance school evaluation in Turkey.

**8. What will happen to information about me that is collected during the study?**
By providing your consent, you are agreeing to me collecting personal information about you for the purposes of this research study. Your information will only be used for the purposes outlined in this Participant Information Statement, unless you consent otherwise. Data management will follow the 2017 General Data Protection Regulation and the University of East Anglia Research Data Management Policy (2015).

Nobody other than me and my supervisors will have access to the data, which will be saved and stored securely on password-protected private computer. The results of this study may be published, but these publications will not contain your name or any identifiable information about you. If you wish, I will forward you the transcript of the interview within two weeks of our meeting and provide feedback about the overall results of this study upon its completion.

**9. What if I would like further information about the study?**
When you have read this information, I will be available to discuss it with you further and answer any questions you may have. If you would like to know more at any stage during the study, please feel free to contact: **Sevda** OZSEZER, PhD researcher: S.Ozsezer@uea.ac.uk, tel. +447447090793.

**10. Will I be told the results of the study?**
You have a right to receive feedback about the overall results of this study. You can tell us that you wish to receive feedback by ticking the relevant box on the consent form. This feedback will be in the form of a one-page lay summary. You will receive this feedback after the study is finished.

**11. What if I have a complaint or any concerns about the study?**
The ethical aspects of this study have been approved under the regulations of the University of East Anglia's School of Education and Lifelong Learning Research Ethics Committee.

If there is a problem, please let me know. You can contact me via the University at the following

address:

*Sevda OZSEZER*
School of Education and Lifelong
Learning University of East Anglia
NORWICH NR4 7TJ
S.Ozsezer@uea.ac.uk

If you would like to speak to someone else, you can contact my supervisors:

*Dr Agnieszka Bates*
Lecturer in Education | School of Education and Lifelong Learning | Lawrence Stenhouse
Building University of East Anglia | Norwich Research Park | Norwich NR4 7TJ
Tel: +44 (0) 1603 592627 | Email: agnieszka.bates@uea.ac.uk

*Professor Nigel F.J. Norris*
University of East Anglia/ Norwich Research Park/ Norwich NR4
7TJ Tel: +44 (0) 1603 592620 | Email: N.Norris@uea.ac.uk

If you are concerned about the way this study is being conducted or you wish to make a complaint
to someone independent from the study, please contact please contact the Head of the School of
Education and Lifelong Learning, Professor Richard Andrews, at Richard.Andrews@uea.ac.uk.

### 12. OK, I want to take part - what do I do next?
You need to fill in one copy of the consent form and return it to me by e-mail
(S.Ozsezer@uea.ac.uk). Please keep the letter, information sheet and the 2$^{nd}$ copy of the consent
form for your information.


**This information sheet is for you to keep**

**CONSENT FORM (1<sup>st</sup> Copy to Researcher)**

I, ................................................................................[PRINT NAME], agree to take part in this research
study.

In giving my consent I state that:

- ✓ I understand the purpose of the study, what I will be asked to do, and any risks/benefits involved.

- ✓ I have read the Participant Information Statement and have been able to discuss my involvement in the study with the researchers if I wished to do so.

- ✓ The researcher has answered any questions that I had about the study, and I am happy with the answers.

- ✓ I understand that being in this study is completely voluntary and I do not have to take part. My decision whether to be in the study will not affect my relationship with the researchers or anyone else at the University of East Anglia now or in the future.

- ✓ I understand that I can withdraw from the study at any time.

- ✓ I understand that I may stop the interview at any time if I do not wish to continue, and that unless I indicate otherwise any recordings will then be erased and the information provided will not be included in the study. I also understand that I may refuse to answer any questions I don't wish to answer.

- ✓ I understand that personal information about me that is collected over the course of this project will be stored securely and will only be used for purposes that I have agreed to. I understand that information about me will only be told to others with my permission, except as required by law.

- ✓ I understand that the results of this study may be published, and that publications will not contain my name or any identifiable information about me.

I consent to:

| ☐ | **Audio-recording** | YES☐ | NO☐ |
| ☐ | **Reviewing transcripts** | YES☐ | NO☐ |

**Would you like to receive feedback about the overall results of this study?**

YES ☐     NO ☐

If you answered **YES**, please indicate your preferred form of feedback and address:

☐ Postal:  _____

☐ Email:  _____

...................................................
Signature/ PRINT name/ Date

**PARTICIPANT CONSENT FORM (2ⁿᵈ Copy to Participant)**

I, .......................................................................................[PRINT NAME], agree to take part in this research
study.

In giving my consent I state that:

✓   I understand the purpose of the study, what I will be asked to do, and any risks/benefits involved.

✓    I have read the Participant Information Statement and have been able to discuss my involvement in the study with the researchers if I wished to do so.

✓    The researcher has answered any questions that I had about the study, and I am happy with the answers.

✓    I understand that being in this study is completely voluntary and I do not have to take part. My decision whether to be in the study will not affect my relationship with the researchers or anyone else at the University of East Anglia now or in the future.
✓    I understand that I can withdraw from the study at any time.

✓    I understand that I may stop the interview at any time if I do not wish to continue, and that unless I indicate otherwise any recordings will then be erased and the information provided will not be included in the study. I also understand that I may refuse to answer any questions I don't wish to answer.

✓    I understand that personal information about me that is collected over the course of this project will be stored securely and will only be used for purposes that I have agreed to. I understand that information about me will only be told to others with my permission, except as required by law.

✓    I understand that the results of this study may be published, and that publications will not contain my name or any identifiable information about me.

I consent to:

| ☐ | **Audio-recording** | YES☐ | NO☐ |
| ☐ | **Reviewing transcripts** | YES☐ | NO☐ |

**Would you like to receive feedback about the overall results of this study?**

YES   ☐         NO      ☐

If you answered YES, please indicate your preferred form of feedback and address:

☐ Postal:        _____

☐ Email:         _____
...............................................................

Signature/ PRINT name/ Date

**Appendix 1b Participant Information Statement and consent form who might be identifiable (however, as explained in Chapter 4, these participants' interview data were redacted to minimise this possibility)**



University of East Anglia

**Sevda** OZSEZER

Research Student
**Science**
School of Education and Lifelong
Learning

**Faculty of Social**

University of East Anglia Norwich Research Park Norwich NR4 7TJ  United Kingdom

Email: S.Ozsezer@uea.ac.uk Tel:  +44 (0)1603 591451

*Approaches to school evaluation in England since 1988*

*PARTICIPANT INFORMATION STATEMENT*

### 1.  What is this study about?

I would like to invite you to take part in my PhD research project about approaches to school evaluation in England since 1988. My research focuses on the changes to school evaluation since the 1988 Education Reform Act (ERA) and their underpinning methodologies to consider whether 'lessons from England' might be relevant to other education systems including the Turkish education system. You are being invited to take part in this research because you are an expert in the field.

This Participant Information Statement tells you about the research study. Knowing what is involved will help you decide if you want to take part in the study. Please read this sheet carefully and ask questions about anything that you don't understand or want to know more about.

Participation in this research study is voluntary. By giving consent to take part in this study you are telling us that you:
- ✓ Understand what you have read.
- ✓ Agree to take part in the research study as outlined below.
- ✓ Agree to the use of your personal information as described.

**Who is running the study?**

Sevda OZSEZER, a PhD student in the School of Education and Lifelong Learning at the University of East Anglia (UEA), is running the study under the supervision of Dr Agnieszka Bates and Professor Nigel Norris.

## 2. What will the study involve for me?

I would like to interview you and, if possible, to audio-record our interview. The interview will seek your views on the changing approaches to school evaluation in England. The interview will be arranged at a time and place convenient for you. I will be happy to meet either in your workplace or a public place such as a café.

## 3. How much of my time will the study take?

The interview will last approximately an hour.

## 4. Do I have to be in the study? Can I withdraw from the study once I've started?

Being in this study is completely voluntary and you do not have to take part. Your decision whether to participate will not affect your current or future relationship with the researchers or anyone else at the University of East Anglia. If you decide to take part in the study and change your mind later, you are free to withdraw at any time up until I have collected the data and analysed the results. You are also free to stop the interview at any time. Should you wish to withdraw from the interview, any recordings will be erased and the information you have provided will not be included in the study results. You may also refuse to answer any questions that you do not wish to answer during the interview.

## 5. Are there any risks or costs associated with being in the study?

Aside from giving up your time, we do not expect that there will be any risks or costs associated with taking part in this study.

## 6. Are there any benefits associated with being in the study?

This study will seek to draw on your knowledge and expertise to enhance school evaluation in Turkey.

## 7. What will happen to information about me that is collected during the study?

By providing your consent, you are agreeing to me collecting personal information about you for the purposes of this research study. Your information will only be used for the purposes outlined in this Participant Information Statement, unless you consent otherwise. Data management will follow the 2017 General Data Protection Regulation and the University of East Anglia Research Data Management Policy (2015).

Nobody other than me and my supervisors will have access to the data, which will be saved and stored securely on password-protected private computer. The results of this study may be published, but these publications will not contain your name or any identifiable information about you. Although every effort will be made to protect your identity, there is a remote risk that you might be identifiable due to your unique role as an expert in the field of school evaluation in England. I will, therefore, forward you the transcript of the interview within two weeks of our meeting if you wish and provide feedback about the overall results of this study upon its completion.

## 8. What if I would like further information about the study?

When you have read this information, I will be available to discuss it with you further and answer any questions you may have. If you would like to know more at any stage during the study, please feel free to contact: **Sevda** OZSEZER, PhD researcher: S.Ozsezer@uea.ac.uk, tel. +447447090793.

## 9. Will I be told the results of the study?

You have a right to receive feedback about the overall results of this study. You can tell us that you wish to receive feedback by ticking the relevant box on the consent form. This feedback will be in the form of a one-page lay summary. You will receive this feedback after the study is finished.

## 10. What if I have a complaint or any concerns about the study?

The ethical aspects of this study have been approved under the regulations of the University of

East Anglia's School of Education and Lifelong Learning Research Ethics Committee.

If there is a problem, please let me know. You can contact me via the University at the following address:

*Sevda OZSEZER*
School of Education and Lifelong Learning
University of East Anglia
NORWICH NR4 7TJ
S.Ozsezer@uea.ac.uk

If you would like to speak to someone else, you can contact my supervisors:

*Dr Agnieszka Bates*
Lecturer in Education | School of Education and Lifelong Learning | Lawrence Stenhouse
Building University of East Anglia | Norwich Research Park | Norwich NR4 7TJ
Tel: +44 (0) 1603 592627 | Email: agnieszka.bates@uea.ac.uk

*Professor Nigel F.J. Norris*
University of East Anglia/ Norwich Research Park/ Norwich NR4
7TJ Tel: +44 (0) 1603 592620 | Email: N.Norris@uea.ac.uk

If you are concerned about the way this study is being conducted or you wish to make a complaint to someone independent from the study, please contact please contact the Head of the School of Education and Lifelong Learning, Professor Richard Andrews, at Richard.Andrews@uea.ac.uk.

### 11. OK, I want to take part – what do I do next?

You need to fill in one copy of the consent form and return it to me by e-mail (S.Ozsezer@uea.ac.uk). Please keep the letter, information sheet and the 2$^{nd}$ copy of the consent form for your information.

**This information sheet is for you to keep**

**PARTICIPANT CONSENT FORM (1ˢᵗ Copy to Researcher)**

I, .......................................................................................[PRINT NAME], agree to take part in this research study.

In giving my consent I state that:

✓ I understand the purpose of the study, what I will be asked to do, and any risks/benefits involved.
✓ I have read the Participant Information Statement and have been able to discuss my involvement in the study with the researchers if I wished to do so.
✓ The researcher has answered any questions that I had about the study, and I am happy with the answers.
✓ I understand that being in this study is completely voluntary and I do not have to take part. My decision whether to be in the study will not affect my relationship with the researchers or anyone else at the University of East Anglia now or in the future.
✓ I understand that I can withdraw from the study at any time.
✓ I understand that I may stop the interview at any time if I do not wish to continue, and that unless I indicate otherwise any recordings will then be erased and the information provided will not be included in the study. I also understand that I may refuse to answer any questions I don't wish to answer.
✓ I understand that personal information about me that is collected over the course of this project will be stored securely and will only be used for purposes that I have agreed to. I understand that information about me will only be told to others with my permission, except as required by law.
✓ I understand that the results of this study may be published. Although every effort will be made to protect my identity, I may be identifiable in these publications due to the nature of the study or results.

I consent to:
☐          **Audio-recording**          YES          ☐          NO          ☐

☐          **Reviewing transcripts**          YES          ☐          NO          ☐

**Would you like to receive feedback about the overall results of this study?**

YES     ☐          NO          ☐

If you answered **YES**, please indicate your preferred form of feedback and address:

☐ Postal:          _____

☐ Email:          _____

.................................................................
Signature/ PRINT name/ Date

**PARTICIPANT CONSENT FORM (2<sup>nd</sup> Copy to Participant)**

I, .................................................................................[PRINT NAME], agree to take part in this research study.

In giving my consent I state that:

- ✓ I understand the purpose of the study, what I will be asked to do, and any risks/benefits involved.
- ✓ I have read the Participant Information Statement and have been able to discuss my involvement in the study with the researchers if I wished to do so.
- ✓ The researcher has answered any questions that I had about the study, and I am happy with the answers.
- ✓ I understand that being in this study is completely voluntary and I do not have to take part. My decision whether to be in the study will not affect my relationship with the researchers or anyone else at the University of East Anglia now or in the future.
- ✓ I understand that I can withdraw from the study at any time.
- ✓ I understand that I may stop the interview at any time if I do not wish to continue, and that unless I indicate otherwise any recordings will then be erased and the information provided will not be included in the study. I also understand that I may refuse to answer any questions I don't wish to answer.
- ✓ I understand that personal information about me that is collected over the course of this project will be stored securely and will only be used for purposes that I have agreed to. I understand that information about me will only be told to others with my permission, except as required by law.
- ✓ I understand that the results of this study may be published. Although every effort will be made to protect my identity, I may be identifiable in these publications due to the nature of the study or results.

I consent to:

| | | | | | |
|---|---|---|---|---|---|
| ☐ | **Audio-recording** | YES | ☐ | NO | ☐ |
| ☐ | **Reviewing transcripts** | YES | ☐ | NO | ☐ |

**Would you like to receive feedback about the overall results of this study?**

YES ☐      NO ☐

If you answered **YES**, please indicate your preferred form of feedback and address:

☐ Postal: _____

☐ Email: _____

.................................................................**Signature/ PRINT name/ Date**

# Appendix 2 Interview questions

1. Please tell me a little about yourself and your role as an educational expert [insert specific role, e.g., Regional Schools Commissioner etc.]?

2. My research focuses on approaches to school evaluation promoted in England since 1988. Would you please tell me what your views on these approaches are?

3. What are your views on Ofsted?

Prompts:

- Ofsted [other agencies relevant to the expert] has been one of the important agencies responsible for school evaluation since 1992. What are your views on Ofsted and the changes to its practices since then?

- In what ways could Ofsted inspection be more effective?

- What if there were no Ofsted inspections?

4. I was wondering about your thoughts on accountability.

Prompts:

- How does it work and what benefits schools, government and other stakeholders can provide?
- In your experience, are there any issues with the current accountability system?
- What would you envisage to be an alternative to accountability?

5. What do you think about high stakes tests, reported in schools' league tables? What are your views of these?

Prompts:

- What are the benefits of this approach?
- What disadvantages can be faced because of this approach?
- What would be an alternative to these?

6. Please tell me what your views are on the parental choice.

Prompts:

- Does parental choice encourage or discourage school improvement?
- Does competition enhance or hider improvement?
- Your perspective upon increasing emphasis on marketisation of successive governments in England

7. School self-evaluation has a long history in England. Over the last four or five decades, self-evaluation has been revisited, embraced, and then forgotten as new ideas and more pressing priorities emerged. I was wondering about your perspective upon these changes.

Prompts:

- Some argue self-evaluation is at the heart of school improvement and some argue it is just a way of self-promotion rather than a process to understanding the weakness and strengths of schools. What do you think about these positions?

- In your experience, have you found that schools feel confident in conducting self-evaluation? (Do you think schools have the capacity to evaluate themselves effectively?)
- (If negative views expressed): What could increase the quality of school self-evaluation?

8. Approaches to school evaluation can be broadly divided into 'qualitative' and 'quantitative'. What are your views on the advantages and limitations of these respective approaches?

9. What would be your suggestions for improving the evaluation of schools?
Prompts:

- What 'lessons' from the approaches to school evaluation in England would, in your view, be helpful in improving school evaluation in other education systems?

10. Is there anything else you would like to add?

# Appendix 3 Interview schedule and data collection

| Category of expertise | Name | Time (21h 18m) | Words (140,356) | Date | Expertise in | Member checking | Redaction |
|---|---|---|---|---|---|---|---|
| Policy Actor | Richard | 01:16:15 | 7,175 | 09/11/18 | Inspector and policy maker-international and the UK | Yes | Not requested |
| | Cooper | 02:03:04 | 11,239 | 12/12/18 | Local Authority management roles with academic experience | Yes | Not requested |
| | Dennis | 00:57:40 | 6,379 | 18/12/18 | Researcher and academic | Yes | Requested and complied |
| | Gabriella | 00:46:18 | 6,080 | 26/02/19 | Inspector (from a devolved system) | Yes | Not requested |
| | Bella | 01:29:49 | 11,162 | 07/06/19 | HMI Inspector, adjudicator with school management experience | Yes | Not requested |
| | Neil | 01:26:00 | 9,502 | 26/06/19 | Policy maker, policy adviser | Yes | Requested and complied |
| Policy Influencer | Nora | 01:19:12 | 11,051 | 25/10/19 | Researcher and academic in the field of assessment | Yes | Not requested |
| | Gordon | 01:18:44 | 9,517 | 18/01/19 | Researcher and academic, specialised in Ofsted inspection, and its alternatives | Yes | Requested and complied |
| | Eduardo | 01:10:58 | 9,279 | 14/06/19 | Policy adviser at Ofsted | No response | N/A |
| | Orion | 02:47:49 | 12,682 | 16/10/19 | Researcher and academic in the field of evaluation | No response | N/A |
| Policy Implementer | Kelvin | 01:49:51 | 13,136 | 10/10/18 | Experienced headteacher | Yes | Requested and complied |
| | Torr | 01:09:23 | 6,926 | 13/03/19 | Senior teacher with academic background | Yes | Not requested |
| | Felicia | 01:04:34 | 8,771 | 28/03/19 | Deputy head with experience in both New Zealand and England | Yes | Not requested |
| | Ned | 01:07:07 | 7,332 | 5/03/19 | Headteacher, lead headteacher, national leader | Yes | Not requested |
| | Kent | 01:32:17 | 10,125 | 17/05/19 | Principal, Ofsted Inspector | Yes | Not requested |

**Table:** Interview time, word count**,** member checking, redaction requests

My name…, and I am a PhD researcher at…, under the supervision of... I have been supported in my study by the Turkish government (Ministry of Education scholarship) in order to research the English approaches to school evaluation and inspection. Upon completion of my PhD, I hope to contribute to the education system in Turkey by working in the Ministry of Education.

I have just read your [the name of Journal] article (Volume [X], Number [Y], [Z]) and found your critique of [a topic/institution] extremely interesting and enlightening. As my research seeks to critically examine approaches to school evaluation in England since 1988 (based on expert interviews), your expertise and knowledge as an expert in the field would greatly enhance my understanding of the changes in the English system and their consequences. I would, therefore, like to invite you to participate in my research. The interview could be face to face or via skype and would take approximately an hour at a time and place convenient to you.

My project has now been approved by the Ethics Committee in the School of Education and Lifelong Learning at the University of East Anglia (UEA) and I attach the Participant Information Statement and interview schedule for your information.

I very much hope that you will be able to take part in my study and look forward to hearing from you.

Kind regards

**Table:** An example interview invitation email ('redacted' for confidentiality)

# Appendix 4 Data analysis

The following is an example of the analytic memos in response to the interview question (3) *What is your views about Ofsted?*

| Interview extracts | Analytic memos pertaining to unintended consequences of Ofsted inspections |
|---|---|
| there are unintended consequences of whatever we [Ofsted] do. (Neil) | unintended consequences are 'natural' |
| You hear stories, I'm sure. Ofsted inspection teams or inspection is going to schools and being a horror story. You know, the terrible things they've said or done or whatever. I've never experienced that… All the inspection teams have been open and transparent and whatever… But those persons whose inspection experience has been positive, you're going to get a reasonable response… And the experience of being inspected is nerve wracking, it's frightening, it's all of those things. But you have to get over that a little bit… I think, we all big enough to cope with that. (Kent) | unintended consequences are an individual (rather than systemic) matter |
| I think, one of the upshots of a lot of developments after 1988s was a reduction in creativity. And also, the whole external pressure which was linked to people losing jobs. So, I was sitting in a meeting in this city with headteachers and then a couple of months later we had another meeting, and someone had disappeared and then another one had disappeared… headteachers were losing their jobs as a direct result of this and the pressures associated with it either through illness or because they were removed or because they got bad Ofsted and then they were removed, or they left before they have to. That's a sick system, it's not a healthy system... it is complete opposite of autonomy, mastery and purpose… (Kelvin) | unintended consequences are systemic ('that's a sick system') and stemming from the prescriptive Ofsted frameworks, punctuative enactments, |
| I worked for a UTC in the last couple of years… Now, that's a specialist school, and intakes in year 10. 98 percent boys and focuses on engineering. That school got placed into 'requires improvement'… because it didn't fit the criteria that that school was never going to fit… It was morale-killing. But that's because they were using an evaluation process that compared us to schools nationally. And we were not a national school. Ofsted works on the basis that you have a nationally representative sample. What school does? Where is that school? (Felicia) | inspections as 'morale-killing' experience - a systemic problem because Ofsted compare schools but not every school has the same context |

**Table:** Example of the analytic memos in response to the interview question (3)

The table below illustrates the codes identified in the interview data under the theme of accountability.

| Theme | Sub-themes | Codes |
|---|---|---|
| Accountability | The essential nature of accountability | Problems with accountability before 1988 |
| | | National accountability to reduce difference between schools |
| | | External accountability as an effective driver of school improvement |
| | | Evolving accountability – changing for the better as it has evolved |
| | Problems with accountability | External 'design' of the official accountability system as a driver of school improvement |
| | | Outcomes-based accountability |
| | | Accountability to whom: accountability to Ofsted |
| | | The punitive enactment of accountability |

**Table:** Nvivo codes

The table below provides example extracts for the alternatives/improvements to Ofsted inspections.

| Data extract | Code for | Theme |
|---|---|---|
| Obviously, we have been for quite a long time now in a period of austerity. Unlike other parts of the education system, we have been affected by that. Our budget has been cut by about 50 percent. That obviously has consequences in terms of things like how often and for how long can we inspect... So, there are compromises that have to be made. (Eduardo) | Improved resources for more comprehensive Ofsted inspection visits | Suggestion to improve Ofsted's inspection methods |
| My biggest bugbear is that non-inspection of outstanding schools. That really aggravates me beyond belief. Because I feel that they should be subject to scheduled inspection as everybody else is. (Kent) | An improved cycle of inspection for 'all' schools | |
| if peer review partnerships are established locally and they are proven to be successful; in other words, children leave school feeling positive about themselves, their futures in the world in which they live; then there is no need for Ofsted. If, however those peer review partnerships break down or if particular schools need specific support, I believe there does need to be some kind of system which means that somebody is able to step in and say: 'we need to work very closely with this school because the children are being let down'. (Ned) | Locally established partnerships-based school evaluation | Reframing the system of school inspections |
| in the current atmosphere, inspection is judgemental, arbitrary and a fault-finding accounting process. However, it needs to be developmental, and seen as educational enterprise and principled. In this regard, re-setting the relationship between Ofsted and the teaching profession on the basis of a two-way educational conversation with schools and teachers is very important. In doing so, both inspectors and teaching professionals have to readjust their mindset… (Dennis) | Non-judgemental inspection with a partnership relation | |

**Table:** Data extracts for the themes: suggestions to improve Ofsted's inspection methods and reframing the system of school inspections