Dr Jenni Ingram, Prof. Pam Sammons and Dr Ariel Lindorff
Oxford University

Observing effective mathematics teaching: a review of the literature

Education Development Trust

UNIVERSITY OF OXFORD

Observing effective mathematics teaching: a review of the literature

Education Development Trust

Education Development Trust Highbridge House, 16–18 Duke Street, Reading, Berkshire RG1 4RU

T +44 (0) 118 902 1000 **E** enquiries@educationdevelopmenttrust.com **W** www.educationdevelopmenttrust.com

ISBN 978-1-909437-97-5

Contents

Welcome to Education Development Trust	**4**
About the authors	**5**
Acknowledgements	**5**
Foreword	**7**
Chapter 1: Introduction	**8**
Different frameworks for classroom observations	9
Breadth of focus	10
Reliability	10
Validity	11
Observation frameworks	12
Chapter 2: Tools for the observation of effective teaching	**14**
The International System for Teacher Observation and Feedback	15
The Quality of Teaching framework	18
The Mathematics Education Traditions of Europe project	21
Chapter 3: The use of frameworks for research in the UK	**24**
Conclusion: Comparing the ISTOF, QoT and METE frameworks	26
Chapter 4: Alternative frameworks	**28**
The Knowledge Quartet	29
Mathematical Quality of Instruction	30
Watson's framework	32
Conclusion	32
Chapter 5: Observing to improve teaching	**34**
Observing effective mathematics teaching	35
Observing to develop teaching	36
Chapter 6: Conclusion	**38**
References	**41**

Welcome to Education Development Trust

At Education Development Trust, we have been improving education around the world for 50 years. We design and implement improvement programmes for school systems, and provide consultancy services deploying specialists internationally.

Our work is informed by our continually refreshed body of research which focuses on the bright spots in education, from education authorities as diverse as those in Vietnam, Kenya, England, New York and Dubai.

Bringing about real change that alters the aspects of a national system that, for many reasons, aren't working so well at the time, requires knowledge and ability to design and implement changes to any of the levers that can impede great educational outcomes. So the ability to affect policy, practices, pedagogy, behaviour, funding, attitudes and more is a prerequisite for a company that can truly claim to transform lives through improving education.

As highly informed agents of change operating in low- to high-income countries with their varying internal contexts, we not only design but also show and enable, so when working with us, everyone involved, from policymakers to school leaders and teachers, is able to apply their new knowledge to drive sustainable system reform.

Our expert knowledge, programme design and implementation expertise is also deployed in delivering Ofsted-rated outstanding careers services in England, and in owning and managing a family of independent schools.

We are a not-for-profit and we are driven by our values of integrity, accountability, excellence and collaboration.

About the authors

Pam Sammons is a Professor of Education at the Department of Education, University of Oxford and a Senior Research Fellow at Jesus College, Oxford. Previously she was a professor at the School of Education, University of Nottingham (2004-2009) and a professor at the Institute of Education University of London (1993- 2004) where she directed the International School Effectiveness & Improvement Centre (ISEIC) 1999-2004. Her research over more than 30 years has focused on school effectiveness and improvement, school leadership, teaching effectiveness and professional development, and promoting equity and inclusion in education.

Jenni Ingram is an Associate Professor of Mathematics Education at the Department of Education, University of Oxford and Vice Principal of Linacre College, Oxford. Previously she was an Assistant Professor at the University of Warwick and a secondary mathematics teacher in inner-city schools in the Midlands. Her research focuses on mathematics education and mathematics teacher education, particularly the role of language in the teaching and learning of mathematics and the using videos in teacher professional development.

Ariel Lindorff is a Research Fellow in the Oxford University Department of Education. She completed her DPhil at the University of Oxford, and was previously a secondary mathematics teacher in inner-city settings in the USA. Her research focuses on educational effectiveness, improvement and equity, often involving advanced quantitative methods and/or mixed methods.

Acknowledgments

Funding for this literature review was provided by the Department for Education (DfE) as part of the TALIS Video Study. The Organisation for Economic Co-operation and Development (OECD) led TALIS Video study, is a pioneering, international study, seeking to improve understanding of which aspects of teaching are related to pupil learning, and the nature of those relationships. The study focuses particularly on the teaching of mathematics in secondary schools. Over approximately one year, more than 750 mathematics teachers in eight countries (including England) will take part in the study. The study will provide a valuable opportunity to explore how teachers teach in different countries and contexts. Education Development Trust, in partnership with Oxford University, manages England's involvement in the study on behalf of DfE. Fieldwork is ongoing, commencing in October 2017 and due to end in July 2018.

Foreword

The quest to understand how best we as educators can teach children and young people is ongoing. We understand a great deal thanks to the work of many scholars and academics who have spent decades studying effective teaching and, in no small way, also to teachers themselves who have been honing their craft over their careers. Despite this, there is still so much we do not know. A recent Organisation for Economic Co-operation and Development (OECD) paper described the challenge:

> *'Supporting teachers has become a top priority across the globe for the improvement of the quality of our education systems. This renewed commitment to the teaching profession is based on evidence that teachers are what makes the greatest difference to learning outside students' backgrounds, and that the quality of our school systems is only as good as the quality of our teachers. A better understanding of what teaching looks like and which approaches are most effective is not a trivial matter. It is critical, for teaching is at the heart of a teacher's role and of the education process.'* [1]

Observation is a tool used in teaching and in educational research, albeit in very different ways in these different disciplines. This review focuses on its use in research and the study of effective teaching. Observation has tremendous power to further our understanding about teaching and learning. The education community is becoming increasingly engaged in the use of video. Simple, commonplace technologies are helping. This review came about as a result of our own engagement in an international study, the pilot OECD Teaching and Learning International Survey (TALIS) Video Study, which uses video to capture and analyse approaches to teaching mathematics. The OECD report goes on to state:

> *'[…] we lack strong evidence about how teaching influences student outcomes and little is based on actual observation of classroom processes […]. Video-based research methods now offer an opportunity to understand what teaching looks like across the globe and, in turn, to enable teachers to learn from their peers […].'* [2]

The tools and approaches that educational researchers employ to understand teaching through observation are a bit of a mystery to many people, particularly non-researchers. This review is a clear and concise descriptive summary of the most tried and tested tools, frameworks and approaches that researchers use to analyse 'observed' teaching. It is an interesting read for anyone involved in the conduct of observation studies linked to effective teaching, particularly where mathematics is an area of focus.

Anna Riggall (PhD)
Head of Research,
Education Development Trust

[1] OECD (2018: 2) [2] Ibid.

Chapter 1

Introduction

This review examines a range of lesson observation frameworks designed for and used in the observation of teaching in mathematics. This includes frameworks specifically designed for international comparisons of teaching practices and teacher effectiveness, as well as those used for teaching development.

Classroom observations can be used in a variety of ways, but they are primarily for the evaluation of teaching, for making comparisons, for professional development or for a combination of these. There is a complex relationship between teaching and learning, and observations of lessons is just one way of examining this relationship. As learning cannot be observed directly, observations usually focus on identifying particular features of the teaching behaviour and the student response. Links are then made with other sources of information, such as student attainment and progress measures, student ratings of teaching or lesson artefacts, such as examples of students' work completed in the lesson.

There is a complex relationship between teaching and learning, and observations of lessons is just one way of examining this relationship

Different frameworks for classroom observations

There is a wide range of observation frameworks available and each is designed to serve a different purpose. The frameworks included in this review are offered as examples of this diversity and to exemplify the issues around the design, usability, validity and reliability of classroom observations. The focus is on frameworks that have focused specifically on the teaching of mathematics but the review also includes two generic frameworks used specifically for international comparisons of the quality and effectiveness of teaching. These two generic frameworks also serve to illustrate the influence of subject matter on the classroom observation process. The review excludes frameworks to evaluate the success of particular interventions or policy initiatives, as these are typically designed to focus on a narrower range of classroom behaviours and are driven by underlying theories of what 'good' teaching might be.

Two key considerations in the design of lesson observation frameworks are the purpose of the observation and who will be conducting it. For systematic observations whose purpose is to identify differences between groups of teachers, for example international comparisons, both validity and reliability are key factors.

Similarly, if the purpose of an observation is to make a judgement about the quality of teaching then reliability is also a key factor. However, as Coe et al.[3] point out, classroom observations that identify teachers as 'above average' or 'below average' are accurate only about 60% of the time. In value-added measures of teacher effectiveness (based on changes in student attainment) the issue of measurement error is acknowledged as important. The use of 'confidence intervals'[4] seeks to address the concept of statistical uncertainty. Measurement error also applies to observation instruments and this is why inter-rater reliability measures are required and observations over more than one lesson are desirable (see Hill, Charalambous and Kraft[5] for an in-depth discussion of the issue of reliability when observing for teacher quality).

Breadth of focus

Another key consideration in the design of lesson observation frameworks is the breadth of focus. Too broad a focus and the framework becomes impractical. The complexity of the classroom means it is unrealistic to try and observe everything in a lesson. Decisions therefore need to be made about what to observe, and these decisions are often framed by the purpose of the schedule or the underlying theoretical basis of what 'good' or 'effective' teaching might be. There is also the potential problem of reducing teaching to a checklist of observable practices, as these cannot take into account the decision-making process behind these behaviours, which are particularly relevant when considering the role of teachers' pedagogic subject knowledge.[6] These decisions may be an essential part of what makes the practices effective and be of particular relevance to mathematics teaching, since teachers' subject knowledge has been found to be more strongly linked to variations in student outcomes than it is in some other subjects.[7]

Reliability

Reliability refers to the extent to which a judgement about a lesson could be replicated. A wide range of factors can affect the reliability of a classroom observation framework score, including the topic being taught, the individual teacher and the observer. However, the extent to which the reliability of an observation framework matters depends on the use of the framework. The issue of reliability is particularly important if observations are used in high-stakes judgements about individual teachers, such as in relation to teacher promotion or pay increases, rather than as a research instrument to establish variation in practice and associations with student outcomes. The Teaching and Learning International Survey (TALIS) Video Study 2018 is not designed to judge teachers but rather to be a source of evidence based research on typical patterns of teaching and their associations with a range of outcomes, including student attainment.

The Teaching and Learning International Survey (TALIS) Video Study 2018 is not designed to judge teachers but rather to be a source of evidence based research on typical patterns of teaching and their associations with a range of outcomes

To improve the reliability of lesson observation frameworks, the focus tends to be on observable behaviours that are 'low inference' – that is, the observer is recording whether something occurs or not without making any judgement about

[3] Coe et al. (2014) [4] 95% confidence intervals mean that, if we were to measure teacher effectiveness 100 times, 95 of the intervals generated would contain the true, unobservable, measure of teacher effectiveness [5] Hill, Charalambous and Kraft (2012) [6] Hewitt (2005) [7] see Hill et al. (2008)

whether the behaviour is 'good' or not. Examples including counting the number of questions asked or the amount of time spent on different activities. 'High-inference' items require the observer to make judgements about what they are observing, such as on the clarity of an explanation or the sequencing of the subject matter.

One key way of improving the reliability of lesson observation scores is to observe more than one lesson with a specific teacher, covering a range of topics and classes. This can help in identifying which variations in practice can be attributed to the topic taught or the class being taught. Another issue with observing single lessons and using these to evaluate teachers is that there is a risk that lessons are performances and not representative of an individual's practice in general or over time. Hill et al.[8] recommend that at least three lessons be observed by at least two different observers to enhance the reliability of observations made.

Another common way of improving reliability is for multiple observers to observe each lesson. Videoing lessons enables the observation of lessons from multiple perspectives and with multiple observers. However, videos do not necessarily capture the full range of what was going on in the lesson, with a single camera frequently focusing only on the teacher. This can mean a loss of information around the interactions between students during the lesson, as well as some interactions between the teacher and the students. Nevertheless, the use of high-quality frameworks and trained observers and pooling the findings of observations of multiple lessons by a range of observers can all help improve the reliability of the observation process.[9]

These issues of reliability are less of a concern when observation is being used as a professional development tool. In these situations, it is the feedback that follows the lesson observation that matters more.[10] Teachers often judge observation feedback to be most useful when it is a subject specialist who can offer advice on how to improve the lesson who conducts it.[11] Evans, Jones and Dawson[12] found that the usefulness of feedback was dependent on whether the observer was a mathematics specialist or not, and that these judgements were based on the advice observers offered on how to improve a lesson. Mathematics specialists offered substantially more suggestions for improvement, with around half of these relating specifically to the subject-centred aspects of the lesson. In addition, teachers have identified peer observation as less threatening and as offering a basis for mutual learning and support to improve practice.[13]

One key way of improving the reliability of lesson observation scores is to observe more than one lesson with a specific teacher, covering a range of topics and classes

Validity

Validity is a particularly challenging issue in the development of lesson observation frameworks, especially given the wide range of purposes for which observations are used. The validity of a framework relates to the extent to which it is measuring what it is intended to measure. Many mathematics-specific schedules focus on the observation of teacher subject knowledge and these frameworks can be correlated with written assessments of this same subject knowledge, although in practice these correlations are not strong. For example, the Mathematical Quality of

[8] Hill, Charalambous and Kraft (2012) [9] Strong, Gargani and Hacifazlioglu (2011) [10] Coe et al. (2014) [11] Wragg et al. (2002) [12] Evans, Jones and Dawson (2014) [13] Muijs and Reynolds (2005)

Instruction (MQI) framework is largely based on this relationship between teacher performances on written assessments of their subject knowledge and observations of these same teachers using these schedules. However, these processes for examining validity are dependent on a range of factors that make them difficult to carry out. To compare the scores you need an existing instrument that was designed to measure the same constructs and that has also been extensively tested for its own validity and reliability. These sorts of studies all need larger sample sizes than are typically associated with observation studies if they are to generate confidence in the statistical analyses.

Lesson observation frameworks can also be validated through comparison with other existing frameworks. If the two frameworks are measuring the same construct, such as the quality of mathematics teaching, then their scores will correlate. However, it can be difficult to find frameworks that measure the same constructs. For example, one study found a low correlation between frameworks specifically focusing on mathematics teaching and more general frameworks.[14] This therefore suggests these frameworks are measuring distinct constructs.

Observation frameworks

This review presents frameworks designed specifically for the comparison of teacher and school effectiveness internationally. The first of these, the International System for Teacher Observation and Feedback (ISTOF),[15] is a general framework examining teacher effectiveness, with over 20 countries, including the UK, involved in its development. This particular framework has been also been used widely for studies of teacher effectiveness within the UK.

The second framework, Quality of Teaching (QoT),[16] was similarly designed to examine primary teaching quality across four countries, including England, and is also used in teacher effectiveness studies within the UK, often alongside the ISTOF.

The final framework was specifically designed for the observation of mathematics lessons, as part of the Mathematics Education Traditions of Europe (METE) project,[17] which involved five European countries, including England. The project also focused on the teaching of three particular topics to students aged ten to 14. The existing research using these three frameworks includes measures of reliability and validity, making them suitable for research into the quality of teaching.

This review also considers three other frameworks, each designed specifically for the observation of mathematics lessons. The first, the Knowledge Quartet (KQ),[18] developed as part of the Subject Knowledge in Mathematics (SKIMA) research programme run by the University of Cambridge,[19] was designed specifically to support the development of mathematics teaching in primary schools in the UK, particularly among student teachers. The basis of this framework is teacher subject knowledge and how it influences the teaching of mathematics. It is now used more broadly by researchers and teacher educators, but remains a tool for teacher professional development.

This review presents three frameworks designed specifically for the comparison of teacher and school effectiveness internationally

[14] Kane and Staiger (2012) [15] Kyriakides et al. (2010) [16] Van de Grift (2007) [17] Andrews (2007) [18] Rowland, Huckstep and Thwaites (2005) [19] See http://www.knowledgequartet.org/introduction/ for further information

A second framework, referred to as the Watson framework, developed in the UK,[20] again was designed with a focus on developing mathematics teaching but this time at the secondary level. This framework starts from the position of identifying what aspects of mathematics are being made available to students through the teaching, rather than focusing on teaching characteristics.

The final framework originates from the US and focuses on the evaluation of mathematics teaching. The Mathematical Quality of Instruction (MQI) framework,[21] has been developed over several years by a team at Harvard, led by Heather Hill, and is now used in a variety of countries to evaluate mathematics teaching. Similarly to the Watson Framework, the focus is on mathematical content and how it is made available to students; similarly to KQ, there is a focus on teacher subject knowledge. However, in contrast with both KQ and the Watson framework, the MQI was designed to provide scores for individual mathematics teachers on a number of discrete dimensions of their mathematics teaching. It is also one of the few mathematics-specific observation frameworks where there has been considerable research examining both its validity and its reliability, both within the US and in other cultural contexts, by means of triangulation of evidence from tests of teacher subject knowledge, observations of practice and 'value-added' measures of student attainment outcomes.[22]

[20] Watson (2007) [21] See https://cepr.harvard.edu/mqi for further information [22] See, for example, Hill, Rowan and Ball (2005); Kane and Staiger (2012)

Chapter 2

Tools for the observation of effective teaching

The International System for Teacher Observation and Feedback was developed by researchers working across 20 countries, including the UK, to specifically explore the effectiveness of teaching internationally.[23]

The International System for Teacher Observation and Feedback

The ISTOF system schedule is based on recording the extent to which the observer agrees that a particular item's description has been observed. The fact that this framework offers feedback for teachers is based on research and expert opinion from more than 20 countries, and that it has been used in a variety of educational contexts, makes it a useful and reliable framework for observing teaching.

This schedule has been used in effectiveness studies within England,[24] combined with the QoT schedule below. The ISTOF has also been used in England in the evaluation of Teach First.[25]

The ISTOF protocol includes 21 indicators grouped into seven components of effective teaching, as table 1, overleaf, shows. There are 45 items in total, which are rated on a five-point Likert scale, from strongly agree (5) to strongly disagree (1), with the option of indicating that it was not possible to observe some features when they were not relevant to or observable in the particular classroom setting.

The ISTOF protocol includes 21 indicators grouped into seven components of effective teaching

[23] Kyriakides et al. (2010) [24] Day et al. (2008); Sammons et al. (2014) [25] Muijs, Chapman and Armstrong (2012)

TABLE 1: ISTOF PROTOCOL [26]

Category	Indicator	Item
Assessment and evaluation	The teacher gives explicit, detailed and constructive feedback	• The teacher makes explicitly clear why an answer is correct or not • The teacher provides his/her feedback on the answers given by the students
	Assessment is aligned with goals and objectives	• Assignments given by the teacher are clearly related to what students learned • The teacher explains how assignments are aligned to the learning goals of the lesson
Differentiation and inclusion	The teacher creates an environment in which all students are involved	• Students communicate frequently with one another on task-oriented issues • Students actively engage in learning
	The teacher takes full account of student differences	• The teacher makes a distinction in the scope of the assignments for different groups of students • The teacher gives additional opportunities for practice to students who need them
Clarity of instruction	The teacher shows good communication skills	• The teacher regularly checks for understanding • The teacher communicates in a clear and understandable manner
	There is clear explanation of purpose	• The teacher clearly explains the purposes of the lesson • The teacher asks students to identify the reasons why specific activities take place in the lesson
	Lessons are well structured	• The teacher presents the lesson with a logical flow that moves from simple to more complex concepts • The teacher implements the lesson smoothly, moving from one stage to another with well-managed transition points
Instructional skills	The teacher is able to engage students	• The teacher provides sufficient wait time and response strategies to involve all types of students • The teacher gives assignments that stimulate all students to active involvement
	The teacher possesses good questioning skills	• The teacher poses questions that encourage thinking and elicit feedback • The length of the pause following questions varies according to the difficulty level of questions (e.g. a question calling for application of abstract principles requires a longer pause than a factual question)
	The teacher uses various teaching methods and strategies	• The teacher uses a variety of instructional strategies during the lesson • The teacher uses different strategies for different groups of students
Promoting active learning and developing metacognitive skills	The teacher helps students develop problem-solving and meta-cognitive strategies	• The teacher invites students to use strategies that can help them solve different types of problems • The teacher invites students to explain the different steps of the problem-solving strategy they are using • The teacher explicitly provides instruction in problem-solving strategies
	The teacher gives students opportunities to be active learners	• The teacher encourages students to ask one another questions and to explain their understanding of topics to one another • The teacher gives students the opportunity to correct their own work
	The teacher fosters critical thinking in students	• The teacher motivates the students to think about the advantages and disadvantages of certain approaches • The teacher asks the students to reflect on the solutions/answers they give to problems or questions • The teacher invites the students to give their personal opinion on certain issues
	The teacher connects material to students' real-world experiences	• The teacher systematically uses material and examples from the students' daily life to illustrate the course content • Students are invited to give their own examples

[26] Adapted from Kyriakides et al. (2010)

TABLE 1: ISTOF PROTOCOL (CONTINUED) [26]

Category	Indicator	Item
Classroom climate	All students are valued	• The teacher demonstrates genuine warmth and empathy towards all students in the classroom • The teacher shows respect for the students in both his/her behaviour and the use of language
	The teacher initiates active interaction and participation	• The teacher creates purposeful activities that engage every student in productive work • The teacher's instruction is interactive (lots of questions and answers)
	The teacher interacts with all students	• The teacher gives turns to and/or involves those students who do not voluntarily participate in classroom activities • The teacher seeks to engage all students in classroom activities
	The teacher communicates high expectations	• The teacher praises students for effort towards realising their potential • The teacher makes clear that all students know that he/she expects their best efforts in the classroom
Classroom management	Learning time is maximised	• The teacher starts the lesson on time • The teacher makes sure students are involved in learning activities until the end of the lesson • Actions are taken to minimise disruption
	Clear rules are evident	• There is clarity about when and how students can get help • There is clarity about what options are available when the students finish their assignments
	Misbehaviour and disruptions are effectively dealt with	• The teacher corrects misbehaviour with measures that fit the seriousness of the misconduct (e.g. s/he does not overact) • The teacher deals with misbehaviour and disruptions by referring to the established rules of the classroom

[26] Adapted from Kyriakides et al. (2010)

The Quality of Teaching framework

The QoT framework was developed by school inspection teams from four countries, including England,[27] to inspect the quality of teaching across these countries in primary schools.[28] This is a value-based framework with high-inference codes requiring the observer to balance the strengths and weaknesses of different features of the classroom practice being observed. The observer awards an overall grade designed to reflect an overall judgement of lesson quality. The initial development of this framework included studies that examined the reliability, inter-rater reliability and validity of the observation framework, specifically focusing on the teaching of mathematics in primary schools. Subsequently, the use of the framework has been extended to other curriculum areas and to secondary lessons within the UK.[29]

The QoT framework requires trained observers to make professional judgements about the practice being observed. It draws on the professional judgement systems used by inspectorates in multiple countries, alongside the educational effectiveness research literature. It has been tested in a number of European countries, which has shown that the measures are reliable and mostly scalar equivalent between different countries.[30]

The framework itself has six quality characteristics, and each item within these includes examples of 'good practice' to improve the reliability of the judgements observers make, as table 2 shows. When the completing the table during a lesson observation, observers must:[31]

- Score each item on a 1–4 scale depending on the balance of strengths and weaknesses. The observer places a teacher on the scale according to the following:

1 = predominantly weak;
2 = more weaknesses than strengths;
3 = more strengths than weaknesses;
4 = predominantly strong.

The observer must score 3 only when all good practice examples (if applicable) are really observed.

- Circle the correct answer:

0 = no, I didn't observe this;
1 = yes, I have observed this.

> The QoT framework was developed by school inspection teams from four countries, including England, to inspect the quality of teaching across these countries in primary schools

[27] Also Belgium, France and the Netherlands [28] Van de Grift (2007) [29] See Day et al. (2008); Sammons et al. (2014) [30] Van de Grift (2013) [31] Van de Grift (2007:148–152)

TABLE 2: QOT FRAMEWORK – LESSON OBSERVATION FORM FOR EVALUATING THE QUALITY OF TEACHING [32]

Rate	Indicators: The teacher...			Observed
Efficient classroom management	• gives a well-structured lesson	1 2 3 4	• ensures clearly recognisable components in the lessons (lesson structure)	0 1
	• ensures the orderly progression of the lesson	1 2 3 4	• ensures entering and leaving the classroom take place in an orderly manner • intervenes in a timely and appropriate way to any order disruptions • acts as a 'watchdog' for agreed codes of behaviour and rules	0 1
	• uses learning time efficiently	1 2 3 4	• ensures there is no loss of time at the start, during or at the end of the lesson • ensures there are no 'dead' moments • ensures the students are not left waiting	0 1
	• ensures efficient classroom management	1 2 3 4	• makes clear which lesson materials should be used • ensures the lesson materials are ready to use • ensures the lesson materials are adapted to the level and experience of the students	0 1
Safe and stimulating learning climate	• ensures a relaxed atmosphere	1 2 3 4	• addresses the children in a positive manner • reacts with humour, and stimulates humour • allows children to make mistakes	0 1
	• promotes mutual respect	1 2 3 4	• encourages students to listen to one another • intervenes when students are being laughed at • takes (cultural) differences and idiosyncrasies into account	0 1
	• supports the self-confidence of students	1 2 3 4	• feeds back on questions and answers from students in a positive way • expresses positive expectations to students about what they are able to take on	0 1
	• shows respect for the students in behaviour and language use	1 2 3 4	• allows students to finish speaking • listens to what students have to say • makes no role-confirming remarks	0 1
	• ensures cohesion	1 2 3 4	• honours the contributions made by students • ensures solidarity between students • ensures events are experienced as group events	0 1
	• stimulates the independence of students	1 2 3 4	• allows students to work independently on another assignment or to take up an individually selected task after completing an assignment • allows students to work with self-correcting materials • has students working on daily and weekly tasks	0 1
	• promotes cooperation between students	1 2 3 4	• provides opportunities for students to help one another • gives assignments that incite cooperation • gives students the opportunity to play together or to carry out assignments together	0 1
Clear instruction	• clarifies the lesson objectives at the start of the lesson	1 2 3 4	• informs students at the start of the lesson about the aims of the lesson • clarifies the aim of the assignment and what the students will learn from it	0 1
	• evaluates whether the objectives have been achieved at the end of the lesson	1 2 3 4	• verifies and/or evaluates whether the aims of the lesson have been achieved • checks the students' achievements	0 1

[32] Adapted from Van de Grift (2007: 148–152)

TABLE 2: QOT FRAMEWORK – LESSON OBSERVATION FORM FOR EVALUATING THE QUALITY OF TEACHING (CONTINUED) [32]

Rate	Indicators: The teacher…			Observed
Clear instruction (continued)	• gives clear instructions and explanations	1 2 3 4	• activates the student's prior knowledge • explains in sequential stages • asks questions that are understood by the students • summarises the lesson materials from time to time	0 1
	• gives clear explanations of the learning materials and the assignments	1 2 3 4	• ensures that every student knows what he/she has to do • clearly indicates the materials that can be used as learning aids	0 1
	• gives feedback to students	1 2 3 4	• checks whether students have understood the lesson materials when he/she is instructing the class • checks whether students are completing the assignments correctly • gives feedback on the way students arrive at their answers • gives feedback on the social functioning involved in the completion of the tasks (group work)	0 1
	• involves all students in the lesson	1 2 3 4	• gives assignments that stimulate students into active involvement • poses questions that initiate reflection • ensures students listen carefully and keep on working • waits sufficiently long to allow students to reflect after posing a question • gives the opportunity to respond to students who don't put their hands up	0 1
	• makes use of teaching methods that activate the students	1 2 3 4	• makes use of conversational forms and discussion forms • provides graduated exercises • permits working in groups/corners • makes use of information and communication technology	0 1
Adaption of teaching	• adapts the instruction to the relevant differences between students	1 2 3 4	• allows students who need less instruction to commence with the work • gives extra instruction to small groups or individual students • does not direct himself exclusively to the middle bracket	0 1
	• adapts the assignments and processing to the relevant differences between students	1 2 3 4	• makes a distinction in the scope of the assignments between individual students • does not give all students the same time to complete the assignment • allows some students to make use of auxiliary materials	0 1
Teaching learning strategies	• ensures the teaching materials are oriented towards transfer	1 2 3 4	• teaches students solution strategies or search and reference strategies • teaches students the use of organisation resources • promotes the conscious use of what has been learned in other (different) areas of learning	0 1
	• stimulates the use of control activities	1 2 3 4	• gives attention to estimatory calculation/anticipatory reading • has solutions related to the context • stimulates the use of alternative solutions	0 1
	• provides interactive instruction and activities	1 2 3 4	• facilitates mutual interaction between students… ensures interaction between pupils and the teacher	0 1

[32] Adapted from Van de Grift (2007: 148–152)

TABLE 2: QOT FRAMEWORK – LESSON OBSERVATION FORM FOR EVALUATING THE QUALITY OF TEACHING (CONTINUED) [32]

Rate	Indicators: The teacher...			Observed
Involvement of pupils	• ensures there is good individual involvement by the pupils	1 2 3 4	• ensures pupils actively listen to the instructions • ensures pupils take part in learning/group discussions • ensures pupils work on the assignment in a concentrated, task-focused way	0 1
Final judgement	**The overall quality of teaching is judged as:**	**1 2 3 4**		

The Mathematics Education Traditions of Europe project

The METE observation framework developed out of a study comparing mathematics teaching in five European countries: England, Finland, Flanders Belgium, Hungary and Spain. The schedule was developed through live observations and then used video recordings of lessons for the main analyses. The lessons focused on the teaching of specific topics with students aged ten to 14: percentages, polygons and linear equations. The main focus of the study was on how mathematics teachers structured students' opportunities for learning.[33] The framework consists of three broad categories, each containing several foci, which were designed to be easily applied across all observers in all countries involved. Each category was designed to be low inference and to address observable behaviours in the lessons. The first category refers to the mathematical foci or observable learning outcomes, as table 3 shows.

TABLE 3: METE MATHEMATICAL FOCI [34]

Mathematical foci	Description – the teacher is seen to emphasise or encourage:
Conceptual	the conceptual development of his or her students
Derivational	the process of developing new mathematical entities from existing knowledge
Structural	the links or connections between different mathematical entities, concepts, properties, etc.
Procedural	the acquisition of skills, procedures, techniques or algorithms
Efficiency	pupils' understanding or acquisition of processes or techniques that develop flexibility, elegance or critical comparison of working
Problem solving	pupils' engagement with the solution of non-trivial or non-routine tasks
Reasoning	pupils' development and articulation of justification and argumentation

The second category for observation focuses on the contexts in which the teachers posed the tasks. It has two dimensions: (1) whether the context was related to the real world or not and (2) whether the data or information used was genuine or invented by the teacher. In this way, the assessment of mathematics classroom activity can be carried out using a two-dimensional grid, as shown in table 4, overleaf.

[32] Adapted from Van de Grift (2007: 148–152) [33] Andrews (2009) [34] Adapted from Andrews (2007: 501)

TABLE 4: METE CONTEXT [35]

Dimensions	Example
The task is explicitly related to the real world and based on data or entities invented by the teacher	• The task of calculating the cost of decorating a hypothetical room is related to the real world but is located in a fantasy of data – the dimensions of the room, the costs of paper, for example. • Revising the cost of a pair of hypothetical trousers after a sale reduction.
The task is explicitly not related to the real world and based on data or entities invented by the teacher	• An invitation to solve the equation x2–3x+1=0 is not based in the real world and the data or entity – the equation itself – is not the product of a student's own activity. • Many text-based questions or exercises would fall into this category.
The task is explicitly related to the real world and based on genuine data or entities	• Testing statistical hypotheses derived from real data collected by students. • Calculating the cost of manufacturing a desk by measuring the desk. It is the act of measurement, which creates genuine data, that feeds back into the real world, as it addresses the cost of making the desks.
The task is explicitly not related to the real world and based on genuine data or entities	• Exploring the minimum value of a quadratic expression of the student's choice has no explicit relation to the real world, but the data – the choice of the individual student – is real. • An invitation to students to measure the length of their desks for no other purpose than to practise the skills of measurement. The task is not explicitly related to the real world because it does not feed back into it, but it is located in real-world, genuine data. In this scenario, the real world provides a background context for the task.

The final category concerns teacher strategies, or 'mathematical didactics', that might be used to facilitate students' learning of mathematics, as table 5 shows. These categories were used both when the teacher was working with a class as a whole and when the students were working individually or as a small group.

TABLE 5: METE TEACHING STRATEGY [36]

Teaching strategy	Description
Activating prior knowledge	Focuses students' attention on mathematical content covered earlier in their careers via a period of revision as preparation for activities to follow.
Exercising prior knowledge	Focuses students' attention on mathematical content covered earlier in their careers via a period of revision unrelated to any activities that follow.
Explaining	Explains an idea or solution. This could include demonstration, explicitly telling or the pedagogic modelling of higher-level thinking. In such instances, the teacher is the informer with little or no student input.
Sharing	Engages students in the sharing of ideas, solutions or answers. This could include class discussions, where the teacher's role is one of manager rather than informer.
Exploring	Engages students in an activity, not teacher directed, from which a new mathematical idea is intended to emerge. This activity could be an investigation or a sequence of structured problems, but in all cases students are expected to articulate their findings.
Coaching	The teacher explicitly offers hints, prompts or feedback to facilitate students' understanding of or ability to perform tasks or to correct misunderstandings.
Assessing or evaluating	Assesses or evaluates students' responses to determine the overall attainment of the class.
Motivating	Addresses students' attitudes, beliefs or emotional responses towards mathematics.
Questioning	The teacher explicitly uses a sequence of questions, perhaps Socratic, so as to lead students to construct new mathematical ideas or clarify or refine existing ones.
Differentiation	Attempts to treat students differently in terms of the kind of activities performed, materials provided and/or the expected outcome to make instruction optimally adapted to the students' characteristics and needs.

[35] Adapted from Andrews (2007) [36] Adapted from Andrews (2007: 503)

Chapter 3

The use of frameworks for research in the UK

The ISTOF, QoT and METE frameworks, have all been used by researchers in the UK to examine teaching, and specifically mathematics teaching, through observations of teaching practice.

These studies often use the frameworks in conjunction with other sources of data, such as teacher interviews and student questionnaires, in order to gain a fuller description of more effective teaching. The findings from these studies relate closely to the existing literature on effective teaching and learning.

The effective classroom practice study used the ISTOF and QoT,[37] alongside teacher questionnaires and interviews, school leader interviews and pupil questionnaires and interviews, to establish a multidimensional picture of effective classroom practice. The frameworks specifically identified several core characteristics of more effective teaching. Specifically, the ISTOF identified clear and coherent lessons with a supportive learning climate; engaging students with assignments and activities; positive classroom management; purposive learning; and quality questioning and feedback for students. The more effective teachers also scored very highly on the QoT characteristics of a supportive lesson climate; proactive lesson management; well-organised lessons with clear objectives; and environmental and teacher support.

> The effective classroom practice study used the ISTOF and QoT, alongside teacher questionnaires and interviews, school leader interviews and pupil questionnaires and interviews

The inspiring teachers study took a mixed-methods approach to characterising inspiring teachers,[38] again using both the ISTOF and QoT, combined with qualitative observations, teacher and school leader interviews, and pupil questionnaires. The sample included 17 teachers representing primary and secondary schools. The comparison between the qualitative observations and the use of the two quantitative frameworks, the ISTOF and QoT, revealed that inspiring teachers also showed strongly the characteristics of more effective teaching. In particular, the teachers identified as inspiring scored particularly highly on the ISTOF in relation to creating a positive classroom climate, classroom management and clarity of instruction. Similarly, these teachers also scored highly on the QoT components related to a safe and orderly school climate, effective classroom layout, clear instruction and effective classroom organisation. As such, while defining inspiring teachers relies more on ideas, such as student engagement and enjoyment, than on effectiveness studies that focus on student academic outcomes, the findings of this study conclude that inspiring teachers are first and foremost highly effective teachers.

The ISTOF has also been used by Muijs, Chapman and Armstrong[39] to explore the effectiveness of Teach First teachers (an alternative teacher certification programme in England). Similarly to the studies above, this framework was used in conjunction

[37] See Day et al. (2008): Kington et al. (2012) [38] Sammons et al. (2014) [39] Muijs, Chapman and Armstrong (2012)

with interviews with teachers and school leaders, and teacher questionnaires. The Teach First teachers demonstrated high levels of the behaviours in the framework that are considered indicators of more effective teachers and they also scored similarly to those teachers observed during the design of the ISTOF framework.[40]

Conclusion: Comparing the ISTOF, QoT and METE frameworks

Whilst the ISTOF and QoT share some similarities in terms of their components and measures, they are both conceptually and practically very different measures of teaching behaviours. Both schedules show sufficient reliable test results for use in the two studies above, but the ISTOF framework scores higher on inter-rater reliability and reliability than the QoT framework. The correlation between teachers' overall scores on the ISTOF and the QoT were strong, positive and statistically significant. Both frameworks provide an overall measure of effective practice but also distinguish different features of practice that can be used when giving feedback to the teachers involved. Similarly, combining these frameworks with field notes can contribute to the usefulness of feedback, as they can provide more detail on student characteristics and prior learning.

The studies that have used these frameworks in combination suggest that there is an overall concept of teacher effectiveness, but also that there are differentiations within this. Consequently, more effective teachers would show both strengths and weakness in particular aspects of their practice that might vary over time, with different topics and with different students. The frameworks themselves identify broad descriptions of more effective practice, but both the effective classroom practice study and the Inspiring Teachers Study found considerable variation in the ways that these broad categories of more effective practice were enacted by the teachers.

The METE framework has been used to explore similarities and differences between teachers, both in England and internationally

The METE framework has been used to explore similarities and differences between teachers, both in England and internationally. It has not been used to compare the effectiveness of particular teaching behaviours but rather what similarities and differences across groups of teachers or across a particular topic, such as linear equations, can tell us about the learning of mathematics. Similarly to the Inspiring Teachers Study, whilst some similarities across the broad categories was observed, there were noticeable differences in the ways that these categories were observed in teachers' classroom practice.[41]

[40] Kyriakides et al. (2010) [41] Andrews, 2009.

Chapter 4

Alternative frameworks

The three frameworks that follow have largely been used in different ways from the three frameworks already discussed.

For example, the Mathematical Quality of Instruction (MQI) framework has been used widely to evaluate the subject knowledge of mathematics teachers, and the number of countries for which it has been validated continues to grow, though there is no study at present showing its validity within the UK. The Knowledge Quartet (KQ) has also been used extensively, but as a professional development tool rather than as a measure of effectiveness or subject knowledge. This contrast mimics the purposes for which these two frameworks were designed: both were initially designed for the observation of primary mathematics lessons but their use has extended to secondary mathematics. The Watson framework was specifically designed for the observation of secondary mathematics classrooms and again has been used predominantly as a professional development tool. The validity and reliability of the MQI has been established in several studies[42] through comparison with written assessments of mathematics teachers' subject knowledge. However, these validity and reliability are not appropriate measures of the usefulness to professional development. This instead relies on how useful teachers and teacher educators have found the frameworks in developing their practice.

The validity and reliability of the MQI has been established in several studies through comparison with written assessments of mathematics teachers' subject knowledge

The Knowledge Quartet

KQ was developed as a framework to support 'productive discussion of mathematics content knowledge between teacher educators, trainees and teacher-mentors'.[43] It was designed as a framework both for lesson observation and for mathematics teaching development. The focus is on mathematics subject knowledge and it is designed to develop both mathematics teaching and mathematics teacher knowledge.

This framework was initially designed through working with primary student teachers and their university tutors and mentors, and through the analysis of videos of teaching on teaching practice, but it is now widely used for professional development purposes at all levels of education.

As table 6 outlines, there are four aspects to KQ. The first category – foundation knowledge – underpins the other categories, as it focuses on the knowledge and beliefs of the teacher, with the other categories focusing on the application of that knowledge in teaching. These categories are not mutually exclusive and episodes within a lesson can be understood in terms of more than one of them. For example, 'a contingent response to a pupil's suggestion might helpfully connect with ideas considered earlier'.[44]

[42] See, for example, Hill et al. (2008) [43] Rowland, Huckstep and Thwaites (2005: 256) [44] Ibid. (259)

TABLE 6: THE KNOWLEDGE QUARTET [45]

Category	Description
Foundation	Propositional knowledge and beliefs concerning: • the meanings and descriptions of relevant mathematical concepts, and the relationships that exist between them; • the different factors that research has shown to be significant in the teaching and learning of mathematics; • the ontological status of mathematics and the purposes of teaching it. Contributory codes: awareness of purpose; identifying errors; overt subject knowledge; theoretical underpinning of pedagogy; use of terminology; use of textbook; reliance on procedures.
Transformation	Knowledge-in-action revealed in deliberation and the choices made in planning and teaching. The teacher transforms and presents his or her own meanings and descriptions in ways designed to enable students to learn. These could include the use of powerful analogies, illustrations, explanations and demonstrations. The choice of examples made by the teacher is especially visible: • for optimal acquisition of mathematical concepts, procedures or essential vocabulary; • for confronting common misconceptions; • for the justification (by generic example) or refutation (by counter-example) of mathematical ideas. Contributory codes: choice of representation; teacher demonstration; choice of examples.
Connection	Knowledge-in-action revealed in deliberation and choice in planning and teaching. Within a single lesson, or across several lessons, the teacher unifies the subject matter and draws out coherence with respect to: • connections between different meanings and descriptions of particular concepts or between alternative ways of representing concepts and conducting procedures; • the relative complexity and cognitive demands of mathematical concepts and procedures, by attention to sequencing of the content. Contributory codes: making connections between procedures; making connections between concepts; anticipation of complexity; decisions about sequencing; recognition of conceptual appropriateness
Contingency	Knowledge-in-interaction revealed through the ability of the teacher to 'think on his/her feet' and respond appropriately to the contributions made by students during a teaching episode. This could be seen in the teacher's willingness to deviate from his/her own agenda, when to develop a student's unanticipated contribution: • might be of special benefit to that pupil; or • might suggest a particularly fruitful avenue of enquiry for others. Contributory codes: responding to children's ideas; use of opportunities; deviation from agenda

Mathematical Quality of Instruction

The MQI was developed by Heather Hill and colleagues at the University of Michigan and Harvard University to reliably measure several dimensions of the work teachers do with students around mathematical content. The MQI is based on a theory of instruction, existing literature on effective instruction in mathematics and an analysis of nearly 250 videotapes of US teachers and teaching. This means that the design is flexible enough to consider the variety of mathematics teaching that occurs in classrooms. The MQI is based on the premise that the mathematical work that occurs in classrooms is distinct from generic features of teaching, such as classroom climate.

[45] Adapted from Rowland, Huckstep and Thwaites (2005: 265–266)

The MQI uses the three key relationships widely used in mathematics education research, often referred to as 'the didactic triangle'.[46] These are the relationships: between the teacher and the mathematics; between the teacher and the students; and between the students and the mathematics, as illustrated in table 7. The framework provides separate teacher scores for five different dimensions, which can each be used to assess these relationships. These dimensions are the richness of the mathematics; errors and imprecision; working with students and mathematics; student participation in meaning making and reasoning; and connections between classroom work and mathematics.

The framework uses video recordings of lessons. Each recorded lesson is then divided into roughly equal length (e.g. 5 or 7.5 minute) segments for scoring by two independent raters. A score is given for each of these five MQI dimensions and the raters also each give the whole lesson an overall MQI score.

TABLE 7: MQI CONSTRUCTS [47]

Construct	Description
Teacher–content relationship	**Richness of the mathematics** Richness includes two pieces: (1) attention to the meaning of mathematical facts and procedures and (2) engagement with mathematical practices and language. Meaning making includes explanations of mathematical ideas and drawing connections among different mathematical ideas or different representations of the same idea. Mathematical practices are represented by multiple solution methods, where more credit is given for comparisons of solution methods for ease or efficiency; by developing mathematical generalisations from examples; and by the fluent and precise use of mathematical language.
	Errors and imprecision This captures whether the teacher makes major errors indicating gaps in mathematical knowledge; whether the teacher distorts content through unclear articulation of concepts; and/or whether there is a lack of clarity in the presentation of content or the launching of tasks.
Teacher–student relationship	**Working with students and mathematics** This investigates whether the teacher accurately interprets and responds to students' mathematical ideas. It also looks at whether the teacher can correct student errors thoroughly, with attention to the specific misunderstandings that led to the errors.
Student–content relationship	**Student participation in meaning making and reasoning** This captures the ways in which students engage with mathematical content, specifically whether students ask questions and reason about mathematics; whether students provide mathematical explanations independently or in response to the teacher's questions; and/or the cognitive requirements of specific tasks, such as whether students are asked to find patterns, draw connections or explain and/or justify their conclusions.
	Connections between classroom work and mathematics This explores whether classroom work has a mathematical point, or whether the bulk of instructional time is spent on activities that do not specifically develop mathematical ideas, for example cutting and pasting or non-productive uses of time, including transitions or discipline.

[46] Chevallard (1985); Brousseau (1997) [47] Adapted from National Center for Teacher Effectiveness (2012)

Watson's framework

Watson's framework[48] is included as it is different in design from the other frameworks discussed in that it 'start[s] from mathematics rather than from teaching'.[49] This framework was designed for use by mathematics teachers, and particularly student teachers, to improve the teaching of mathematics. Again, the framework focuses on observable teacher behaviours but also includes considerations of the 'kinds of shift a learner might be hoped to make during mathematical activity'.[50] Table 8 presents the 'dimensions of mathematical pedagogic orientation', the relevant tasks or prompts that are observable and the shifts required for each dimension.

Conclusion

These three frameworks all focus on the particular features of mathematics teaching and enable us to observe the mathematical content and its presentation, as well as some of the more general features focused on in the earlier frameworks. Each of these frameworks have been designed for a specific purpose, such as the subject knowledge of mathematics teachers or as a professional development tool, as well as for primary teaching or secondary teaching. Whilst their use in other settings or for other purposes has not been validated by research, many teachers are finding them a useful framework for analysing their own teaching.

Frameworks have been designed for a specific purpose, such as the subject knowledge of mathematics teachers or as a professional development tool, as well as for primary teaching or secondary teaching

[48] Watson (2007) [49] Ibid. (118) [50] Ibid. (120)

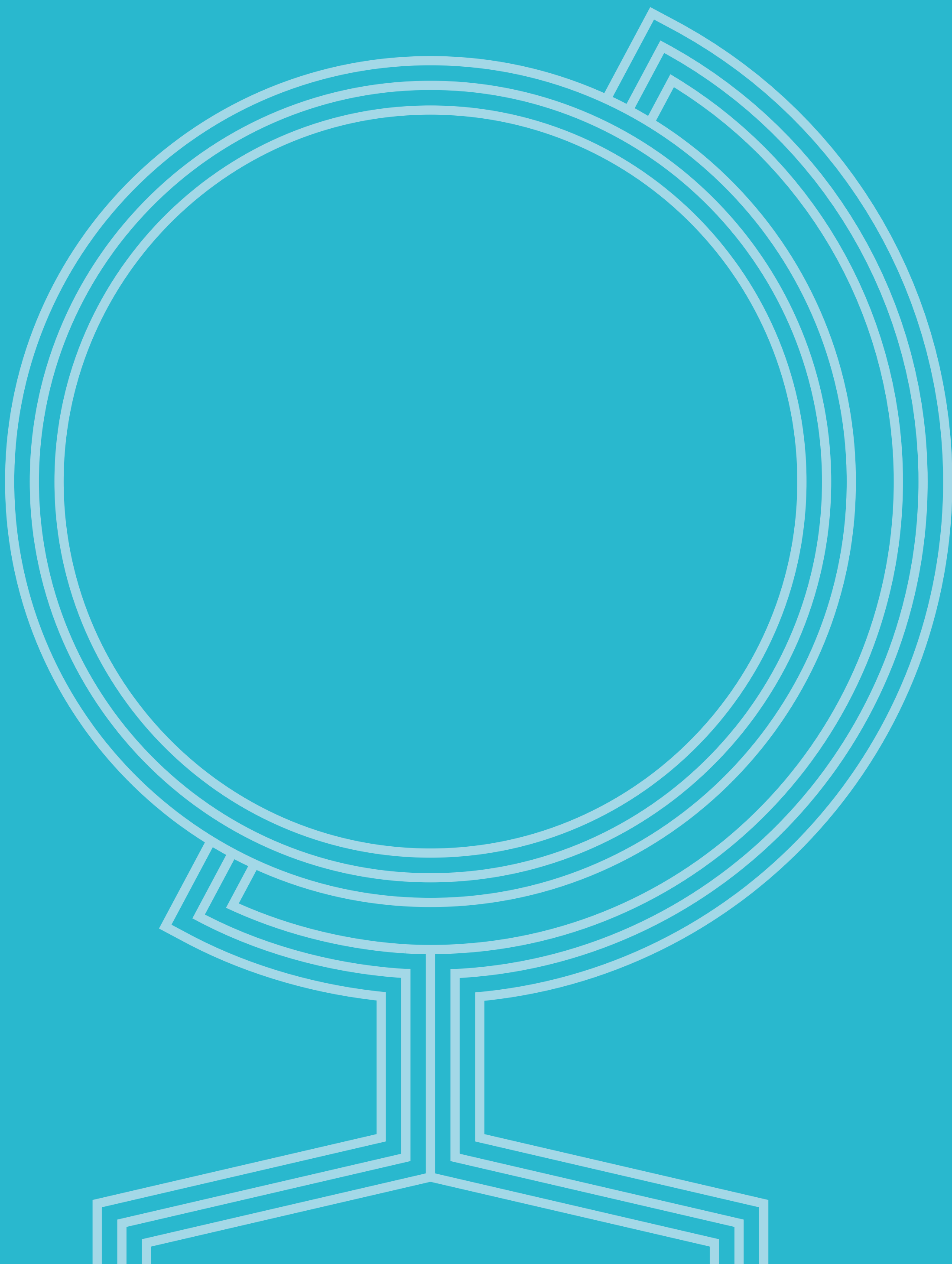
TABLE 8: WATSON'S FRAMEWORK [51]

Dimensions	Tasks/prompts	Shifts
Teacher makes or elicits declarative/nominal/factual/ technical statements	• Say what the lesson is about • Information giving • Define terms • Tell/know/ask facts, definitions and techniques • 'Research' facts definitions, and techniques	Remember
Students are expected to…	• Imitate method, copy object • Follow procedure • Find answer using procedure • Give answers	Fluency, report/record actions
Teacher directs student perception/attention	• Tell/show objects that are perceived as having a single feature • Tell/show objects that are perceived as having multiple features • Tell/show multiple objects • Indicate identification of characteristics/properties • Indicate classification • Indicate comparison • Indicate identification of variables and variation • Summarise what has been done	Public orientation towards concepts, methods, properties, relationships
Teacher asks for student response	• Tell what to think about • Use prior knowledge • Find answer without known procedure • Visualise • Seek pattern • Compare, classify, describe • Explore variation • Informal induction • Informal deduction • Create objects with one feature • Create objects with multiple features • Exemplify • Express in 'own' words	Personal orientation towards concepts, methods, properties, relationships
Discuss implications	• Varying the variables deliberately • Adapting procedures • Identifying relationships • Explication/justification • Induction/prediction • Deduction	Analysis, focus on outcomes and relationships
Integrate and connect mathematical ideas	• Clarify • Associate ideas • Generalisation • Re-description • Summarise development of ideas • Abstraction • Objectification • Formalisation • New definition	Synthesis, connection
Affirm or act as if we know	• Explore properties of new objects • Adapt/transform ideas • Application to more complex mathematics • Application to other contexts • Evaluation for development of new idea • Prove	Rigour, objectification, use

[51] Adapted from Watson (2007: 120)

Chapter 5

Observing to improve teaching



Classroom observation frameworks developed for international comparisons need to consider cultural factors that may influence both the reliability and the validity of the framework.

Teaching characteristics that matter to one country but not another can skew reliability if observers from a particular country consistently score a particular characteristic highly or consistently low. For example, if it is specific practice within a country to share the learning objectives for the lesson on the whiteboard at the beginning of the lesson then any item that records whether this occurred or not will generally score highly within this country, therefore giving the impression of a high degree of reliability between observers for this item. Similarly, some characteristics that matter particularly to one country but not others may be omitted entirely from the framework and thus not necessarily measure some features of teaching in that country.

Issues related to individual items within observation frameworks also arise, with specific characteristics potentially having different meanings within different cultures. Andrews[52] gives the example of the meaning of the concept of 'realistic context' varying considerably across countries. Additionally, the mathematics education literature widely debates issues around variation in the meaning of the terms 'problem', 'problem solving', 'reasoning' and 'cognitive demand'.[53]

Observation schedules designed for international comparisons need to take into account the possibility that some items may not be deemed relevant in some countries, as well as the need to identify actual differences in teacher practice across countries. The frameworks also need to address the specific policy interests or research questions of each of the countries involved at the time. Using expert opinion from mathematics educators and teachers in different countries can help focus observations on areas where there is greater agreement in the definition of concepts to enhance the validity of any instruments used.

Observation schedules designed for international comparisons need to take into account the possibility that some items may not be deemed relevant in some countries

Observing effective mathematics teaching

Observation frameworks can be used to characterise and identify both effective teaching and effective teachers. Research into what constitutes effective teaching is a growing field internationally, and aims to identify characteristics that can then be used as a basis for developing practice. The frameworks above are all designed to focus on teaching rather than teachers, and all drew on the existing research literature on effective teaching in their design. Observations of particular lessons may not give a particularly accurate impression of a teacher's practice, as is shown

[52] Andrews (2007) [53] Polya (1965); Schoenfeld (1989)

in studies that look at the variance in scores explained by different factors, such as the topic taught or the nature of the class.[54] Therefore, although the frameworks may give a reliable and valid judgement about the quality of a lesson, they cannot on their own be used to make judgements about whether a particular teacher is effective or not.

Different frameworks characterise effectiveness in a variety of ways. The most common measure in effectiveness studies relates to student gains, usually in terms of value-added assessment scores. Other measures used include attainment levels, measures of student engagement and measures of student attitudes or beliefs.[55] The number of studies relating observation scores to student achievement gains is growing[56] but, although many of the findings are significant, the correlations are still relatively low. The MQI framework in particular has been compared to student achievement gains on a variety of mathematics assessments, as well as teacher subject knowledge assessments, though the relationship with teacher subject knowledge assessments is stronger than the relationship with student achievement gains in mathematics.

The most common measure in effectiveness studies relates to student gains, usually in terms of value-added assessment scores

In England, two key reports on effective teaching have recently been produced, by Coe et al.[57] and Ko, Sammons and Bakkum.[58] Both of these reports synthesise a range of research that has included mathematics teachers, and identify a range of characteristics of effective teaching. The majority of these characteristics are observable, but the extent to which the frameworks described include them varies. The general frameworks in particular include a wider range of these characteristics in their schedules but rarely go beyond stating whether this was a feature of the lesson or not. The subject-specific frameworks have a far narrower focus but also explore different features within the characteristics to give a finer-grained account of specific practices.

Pedagogic subject knowledge is a dominant feature of those frameworks specifically designed for mathematics teaching and is also emphasised in the Ofsted inspections within England.[59] Mathematics is one of the few subject areas where there is research evidence that the subject knowledge of teachers affects students' learning.[60] However, this relationship is between the nature of a teacher's subject knowledge rather than any qualifications in mathematics.[61] The KQ and MQI frameworks both focus on observing the nature of subject knowledge as evidenced in practice.

Observing to develop teaching

Many of the frameworks described above are also designed to give meaningful feedback to the teachers being observed. Muijs and Reynolds[62] have provided guidance on appropriate ways to use observation to support professional development. Coe et al.[63] emphasise the role of observations and the importance of quality feedback in teacher development, particularly given the issues around the reliability of observations. The reliability of many of the observations that occur within English schools at present is a concern, and can amount to judgements

[54] Kane and Staiger (2012) [55] Sammons, Lindorff and Kington (2016) [56] e.g. Day et al. (2008); Kane and Staiger (2012) [57] Coe et al. (2014) [58] Ko, Sammons and Bakkum (2013) [59] Ofsted (2008) [60] Hill, Rowan and Ball (2005) [61] Askew et al. (1997) [62] Muijs and Reynolds (2005) [63] Coe et al. (2014: 3)

made by tossing a coin, according to Coe et al., who encourage considerable caution when interpreting judgements made. The simplest ways of improving this situation entail using observations only for low-stakes purposes; ensuring those undertaking the observations receive appropriate training; and using observers with some subject-specific expertise.

There are also many issues involved in using observations and the accompanying feedback to support teacher development. Some characteristics of teaching that matter most in professional development are more challenging to measure reliably – for example the quality of a mathematical explanation.[64] Many characteristics that are the focus of subject-specific observation frameworks require a certain amount of subject expertise on the part of the observer. Even low-inference codes, such as whether multiple representations were used, include knowledge of what does or does not count as a different representation within a mathematical context. This points to the need to use observers with appropriate mathematical knowledge and to have appropriate and rigorous observation training and reliability checks.

For professional development purposes it is the quality of the feedback that matters the most. Evans, Jones and Dawson[65] showed that the perceived usefulness of feedback was dependent on whether the observer was a mathematics specialist or not, and that these judgements were based on the advice observers offered on how to improve a lesson. Mathematics specialists offered substantially more suggestions for improvement, both in relation to the mathematics-specific aspects of the lesson and in terms of more general features of teaching. However, teacher shortages have also been shown to have a negative effect on giving meaningful feedback based on classroom observations.[66] This is particularly relevant to the observation of mathematics teachers within England.

For professional development purposes it is the quality of the feedback that matters the most

Higher-stakes observations can incentivise particular practices, such as sharing learning objectives on the board at the beginning of a lesson. However, even though the mere inclusion of these specific practices may improve the quality of teaching students are experiencing, these practices may not be applied in the way the research has shown makes them effective. Providing feedback based on observations can allow teachers to perform these practices more effectively.

[64] Hill et al. (2008) [65] Evans, Jones and Dawson (2014) [66] Reynolds, Muijs and Treharne (2003)

Chapter 6

Conclusion

This review of existing classroom observation practices demonstrates that classroom observation frameworks can be designed and used for a variety of purposes.

The review has drawn on six specific frameworks. These vary in terms of whether they are mathematics-specific or not; whether their purpose is to identify the effectiveness or quality of teaching, or primarily for professional development; and the extent to which there have been studies into their reliability and validity.

Important points to consider regarding frameworks for classroom observations include the following:

- Observations on their own are not reliable enough to make secure judgements about the quality of individual lessons for the purposes of 'high-stakes' judgements. Nonetheless, reliability can be improved through the use of multiple trained observers, observing a range of lessons with the same teacher, using more than one instrument; and through combining the observation data with other sources of evidence, such as student questionnaires, attainment-based measures of value-added and examples of their work during the lesson.
- In international studies, it is important to recognise that expert opinion may be needed to offer a clear focus for any lesson observations conducted; to ensure agreement about the main constructs to be covered and how they are measured; and to provide training to enhance the reliability and validity of the data to be used in any cross-country comparisons.

Observations on their own are not reliable enough to make secure judgements about the quality of individual lessons for the purposes of 'high-stakes' judgements

References

Andrews, P. (2007). 'Negotiating meaning in cross-national studies of mathematics teaching; kissing frogs to find princes', *Comparative Education,* 43 (4), 489–509.

Andrews, P. (2009). 'Comparative studies of mathematics teachers' observable learning objectives: validating low inference codes', *Educational Studies in Mathematics,* 71, 97–122.

Askew, M., Rhodes, V., Brown, M., Wiliam, D. and Johnson, D. (1997). 'Effective teachers of numeracy'. Report of a study carried out for the Teacher Training Agency. London. London: King's College, University of London.

Brousseau, G. (1997). *Theory of didactical situations in mathematics.* Edited and translated by N. Balacheff, M. Cooper, R. Sutherland and V. Warfield. Dordrect: Kluwer.

Chevallard, Y. (1985). *La transposition didactique – du savoir savant au savoir enseigné.* Grenoble: La Pensée Sauvage.

Coe, R., Aloisi, C., Higgins, S. and Major, L. E. (2014). *What makes great teaching? Review of the underpinning research.* London: Sutton Trust.

Day, C., Sammons, P., Kington, A., Regan, E., Ko, J., Brown, E., Gunraj, J. and Robertson, D. (2008). 'Effective classroom practice: a mixed method study of influences and outcomes'. Full Research Report. ESRC End of Award Report. Swindon: ESRC.

Evans, S., Jones, I. and Dawson, C. (2014). 'Do subject specialists produce more useful feedback than non-specialists when observing mathematics lessons?' *Proceedings of the 38th Conference of the International Group for the Psychology of Mathematics Education,* 3, 33–40.

Hewitt, D. (2005). 'Talking about subject-specific pedagogy', *For the Learning of Mathematics,* 25 (3), 32–36.

Hill, H. C., Rowan, B. and Ball, D, L. (2005). 'Effects of teachers' mathematical knowledge for teaching on student achievement', *American Educational Research Journal,* 42 (2), 371–406.

Hill, H. C., Blunk, M., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L. and Loewenberg Ball, D. (2008). 'Mathematical knowledge for teaching and the mathematical quality of instruction: an exploratory study', *Cognition and Instruction,* 26 (4), 430–511.

Hill, H. C., Charalambous, C.Y. and Kraft, M. A. (2012). 'When rater reliability is not enough: teacher observation systems and a case for the generalizability study', *Educational Researcher,* 41 (2), 56–64.

Kane, T. J. and Staiger, D. O. (2012). *Gathering feedback for teaching: combining high-quality observations with student surveys and achievement gains.* Seattle, WA: Bill & Melinda Gates Foundation.

Kington, A., Regan, E., Sammons, P. and Day, C. (2012). 'Effective classroom practice: a mixed-method study of influences and outcomes'. Research Paper. Nottingham: The Nottingham Jubilee Press.

Ko, J., Sammons, P. and Bakkum, L. (2013). *Effective teaching: a review of research and evidence.* Reading: CfBT Education Trust.

Kyriakides, L., Creemers, B. P. M., & Teddlie, C. (2010). The international system for teacher observation and feedback. In: P.L. Peterson, E. L. Baker, & B. McGaw (Eds.), International encyclopedia of education, 3rd ed., Vol. 3. Oxford, England: Elsevier.

Muijs, D., Chapman, C. and Armstrong, P. (2012). 'Teach First: pedagogy and outcomes. The impact of an alternative certification programme', Journal for Educational Research Online, 4 (2), 29–64.

Muijs, D. & Reynolds, D. (2005). Effective teaching: Evidence and practice. London: Sage.

National Center for Teacher Effectiveness. (2012). *Measuring MQI* **[http://isites.harvard.edu/icb/icb.do?keyword=mqi_training&pageid=icb.page394700]** (accessed 19 July 2016).

OECD (Organisation for Economic Co-operation and Development). (2018). *What does teaching look like: a new video study.* Teaching in Focus #20: Paris: OECD.

Ofsted. (2008). *Mathematics: understanding the score.* London: Ofsted.

Polya, G. (1965). *Mathematical discovery: on understanding, learning, and teaching problem solving (Vol. II).* New York: John Wiley and Sons.

Reynolds, D., Muijs, D. and Treharne, D. (2003). 'Teacher evaluation and teacher effectiveness in the United Kingdom', Journal of Personnel Evaluation in Education, 17, 83–100.

Rowland, T., Huckstep, P. and Thwaites, A. (2005). 'Elementary teachers' mathematics subject knowledge: the knowledge quartet and the case of Naomi', *Journal of Mathematics Teacher Education,* 8, 255–281.

Sammons, P., Kington, A., Lindorff-Vijayendran, A. and Ortega, L. (2014). *Inspiring teachers: perspectives and practices.* Reading: Education Development Trust.

Sammons, P., Lindorff, A. and Kington, A. (2016). 'Inspiring teaching: learning from exemplary practitioners', *Journal of Professional Capital and Community,* 1 (2), 124–144.

Schoenfeld, A. H. (1989). 'Teaching mathematical thinking and problem solving'. In: L. B. Resnick and L. E. Klopfer, eds, *Toward the thinking curriculum: current cognitive research.* Alexandria, VA: Association for Supervision and Curriculum Development.

Strong, M., Gargani, J. and Hacifazlioglu, O. (2011). 'Do we know a successful teacher when we see one? Experiments in the identification of effective teachers', *Journal of Teacher Education,* 62 (4), 367–382.

Teddlie, C. Creemers, B. Kyriakides, L. Muijs, D. & Fen, Y. (2006) The international system for Teacher Observation and Feedback: Evolution of an international study of teacher effectiveness constructs, Educational Research and Evaluation, 12:6, 561-582, DOI: HYPERLINK **https://doi.org/10.1080/13803610600874067**

Van de Grift, W. (2007). 'Quality of teaching in four European countries: a review of the literature and application of an assessment instrument', *Educational Research,* 49 (2), 127–152.

Van de Grift, W. (2013). 'Measuring teaching quality in several European countries', *School Effectiveness and School Improvement,* 25 (3), 295–311.

Watson, A. (2007). 'The nature of participation afforded by tasks, questions and prompts in mathematics classrooms', *Research in Mathematics Education,* 9 (1), 111–126.

Wragg, E. C., Wilkley, F., Wragg, C. M. and Haynes, G. S. (2002). *Teacher appraisal observed.* London: Routledge.

Education Development Trust… we've changed from CfBT

We changed our name from CfBT Education Trust in January 2016. Our aim is to transform lives by improving education around the world and to help achieve this, we work in different ways in many locations.

CfBT was established nearly 50 years ago; since then our work has naturally diversified and intensified and so today, the name CfBT (which used to stand for Centre for British Teachers) is not representative of who we are or what we do. We believe that our new company name, Education Development Trust – while it is a signature, not an autobiography – better represents both what we do and, as a not for profit organisation strongly guided by our core values, the outcomes we want for young people around the world.

Education Development Trust

Education Development Trust Highbridge House, 16–18 Duke Street, Reading, Berkshire RG1 4RU

T +44 (0) 118 902 1000 **E** enquiries@educationdevelopmenttrust.com **W** www.educationdevelopmenttrust.com