# BRITISH COUNCIL

# TeachingEnglish

# Teacher evaluation: Global perspectives and their implications for English language teaching
## A literature review

Simon Borg

# Teacher evaluation: Global perspectives and their implications for English language teaching
## A literature review

Simon Borg

**About the author**
Simon Borg has been involved in English language teaching for 30 years, working as a teacher, teacher educator, lecturer, researcher and consultant in a range of international contexts. Simon is recognised internationally for his work on teacher cognition, professional development and teacher research. He works as an ELT consultant, with a particular focus on designing, facilitating and evaluating language teacher professional development programmes. Full details of his work are available at simon-borg.co.uk

**Cover photo:** © Mat Wright

# Contents

# Acknowledgements

# Foreword

The purpose of our TeachingEnglish series of publications is to improve the quality of English learning and teaching worldwide. When the British Council works with teachers, teacher educators, institutions and educational authorities internationally, we emphasise some underlying values and principles that we believe apply to all educational improvement programmes. One principle is to ensure the involvement of all stakeholders in the process, which will normally include at least policy makers, researchers, school or college leaders, in addition to teachers, parents and learners. If one group of stakeholders is not taken into account, the success of the improvement programme is at risk. A second principle is to take all the elements of the educational offer into consideration in the planning and implementation of educational change. We see the four most fundamental elements as being curriculum, learner assessment, teaching, and overall quality assurance. Again, if one of these elements is neglected the whole reform project can be jeopardised.

This report fits squarely into this model of the design and implementation of education improvement and reform. It looks in particular at quality assurance of the teaching element of the educational offer. It is in the area of teaching quality improvement that most resource is spent in reform projects, and with good reason: the quality of teaching has been shown to be the most important influencer of learner success in schools. We believe that teachers are professionals, and that all professionals should be committed to lifelong learning and continuing professional development. But teachers, their schools, universities and their educational authorities cannot design effective teacher improvement programmes without knowing the levels of quality that teachers are operating at in different professional practices. When this is known, the authority or the individual teacher can start to decide where best to focus in the teacher development and education programmes. The best performing education systems invest in the evaluation of teacher performance. This study gives an invaluable overview of the field, covering and critiquing a range of methodologies including classroom observation, professional conversations, learner outcomes, teacher tests and portfolios, as well as self, peer, student, and parent evaluation.

Simon Borg is a leading researcher and thinker on teacher identity, evaluation, and development, and has wide experience of educational systems internationally. While the prime focus of Professor Borg's work is language teaching, particularly English, the lessons learned and recommendations here apply just as well to the evaluation of teachers and teaching in all subjects and at all levels. Much of the literature reviewed has come from educational contexts far wider than just English language teaching. We believe that this short study will be a valuable resource to educational policy makers and managers who want to make valid and reliable evaluations of the teaching and teachers within their institutions and systems. The British Council thanks Simon Borg and all those who helped complete this study and is pleased to share the findings with the international education community.

**John Knagg OBE FAcSS**
Global Head, English for Education Systems
British Council

# Executive summary

The quality of education that students receive is influenced by several factors, but key among these is the quality of their teachers. A significant concern for educational institutions and organisations, therefore, is how to assess and improve the quality of the teachers they employ. With particular attention to state education systems, this review draws on an international body of educational literature to highlight different ways in which teacher evaluation can take place and to make recommendations for teacher evaluation in English language teaching (ELT).

Several key messages emerge from this review.

1   While the literature on teacher evaluation is vast, limited evidence is available of how teacher evaluation actually takes place in state school ELT around the world and more research, including case studies in specific contexts, is required.

2   Although classroom observations remain the most common way of evaluating teachers worldwide, several other strategies for teacher evaluation are available. These include analysing student outcomes, teaching portfolios, student evaluations of teaching, self-evaluation, teacher tests, professional conversations, peer evaluation and parent feedback.

3   Teacher evaluation is a multi-dimensional process and it should draw on multiple sources of evidence collected over time. No one source of evidence can provide sufficiently broad and accurate insights into teacher quality.

4   Decisions about how teachers are evaluated must be aligned with other components of a broader teacher evaluation framework, such as the purposes and focus of teacher evaluation and the resources available to implement it.

5   While teacher evaluation can serve summative and formative purposes, an effective teacher evaluation system will generate results that are used to support the professional development of teachers and, consequently, to improve student learning.

6   Teacher evaluation is facilitated when the criteria against which teachers will be evaluated are clearly defined. These criteria should be aligned with the teaching standards prevalent in the context where teachers work.

7   Two factors which contribute significantly to the quality of teacher evaluation are well-designed evaluation tools and rigorous evaluator training. Shortcomings in these aspects of teacher evaluation significantly impair the quality of the process.

8   Teacher quality is multi-faceted; it includes not just teacher performance in the classroom and its impact on students, but broader aspects of teachers' work such as planning, reflection, professional development, contributions to school leadership and collegiality. Teacher evaluation should reflect this extended view of teacher quality.

9   Teachers and other stakeholders in the process are more likely to respond positively to teacher evaluation when they are involved in decisions about its design and implementation.

10   The effective implementation of teacher evaluation can be hindered by a range of technical, financial, human resource, political, professional and social factors. It is essential that these are given due attention when decisions about teacher evaluation are being made.

# 1

# Introduction

The key question that motivated this review is a straightforward one: how can the quality of English language teachers be evaluated? Teacher evaluation is not in itself a new concept, but it has in recent years become a major focus for policy, research and practical activity in education. Two factors have contributed to this intense contemporary interest in teacher evaluation. The first is the recognition that teachers are the most important school-level influence on student achievement (OECD, 2013a).[1] Teacher quality matters, and a concern for teacher evaluation has become an important element in teacher policies generally (World Bank, 2012). Secondly, there is increasing evidence that conventional approaches to teacher evaluation are not effective. For example, while it is acknowledged that teacher evaluation should differentiate among teachers (Glazerman et al., 2011), evidence from the USA has shown that this is often not the case and that most teachers are rated as being highly competent (Kraft and Gilmour, 2017; Weisberg, Sexton, Mulhern and Keeling, 2009). More generally, Darling-Hammond (2013) has highlighted several common problems in teacher evaluation systems, including a lack of consistent standards of good practice, inadequate time for evaluation to take place, and the fact that teacher evaluations do not support teacher development.

As reflected extensively in the recent literature, effective teacher evaluation is now recognised as a key component of successful education systems. This literature will be discussed here with the aim of arriving at some recommendations for the evaluation of English language teachers in state schools. One point to note from the outset, though, is that literature which focuses specifically on teacher evaluation in English language teaching (ELT) in primary and secondary education is limited. Texts such as *Evaluating teacher effectiveness in ESL/EFL contexts* (Coombe, Al-Hamly, Davidson and Troudi, 2007) and *Teacher evaluation in second language education* (Howard and Donaghue, 2015) do include some material of relevance but their overall focus is on higher education systems. Much information about national teacher evaluation frameworks (for example, such as those that apply in Chile or Korea) is available and it will be assumed here that such frameworks are relevant to all subjects including ELT. Nonetheless, it is clear that there is a genuine lack of research into how ELT teacher evaluation in primary and secondary schools around the world actually takes place, and one key conclusion from this report is that much more work of this kind is necessary to ascertain what happens in practice, how distinct it is or needs to be compared to teacher evaluation in other subjects and how

it is perceived by teachers (there is evidence that teachers may not respond positively to teacher evaluation reform even when it seems to reflect good practice as defined in the literature – for examples from Japan and Korea see respectively Katsuno, 2016; Youngs, Kim and Pippin, 2015.

Although the primary focus here is on the specific strategies that can be used to evaluate teachers, these strategies must be seen as part of a broader system of teacher evaluation. I begin, therefore, by clarifying some terminology that is common in the teacher evaluation literature before outlining key features of teacher evaluation systems. This will be followed by the core section on teacher evaluation strategies. The final part of the report discusses challenges that can arise when teacher evaluation systems are being implemented. In conclusion, several recommendations are made which draw on the literature discussed here and which provide guidance for organisations and education systems seeking to improve the effectiveness with which teachers of English are evaluated.

---

[1] The Organisation for Economic Co-operation and Development (OECD) is a forum of 34 industrialised countries that develops and promotes economic and social policies – www.oecd.org

# 2

# Key concepts

Various terms are used in the literature to describe the process of assessing teacher quality. **Teacher evaluation** is the most common and is the primary term that is used in this review; OECD (2015, p. 502) defines it as 'the evaluation of individual teachers to make a judgement about their work and performance using objective criteria'. **Teacher appraisal** is sometimes used interchangeably with teacher evaluation, but it more specifically refers to 'formal performance reviews, usually conducted by a school level supervisor, to judge individual teacher performance' (Looney, 2011, p. 442). In this sense, teacher appraisal is one aspect of the broader process of teacher evaluation, which may also include informal assessments. **Teacher assessment** is also used in discussions of teacher evaluation (Haertel, 1991; Shulman, 1988), typically to refer to the measurement of specific domains of teacher quality such as teacher knowledge or performance in the classroom. In this sense teacher assessment contributes to the broader process of teacher evaluation (see the discussion of teacher assessment and evaluation systems in National Education Association, 2010).[2]

Another term in the literature is **teacher effectiveness**, as seen in the titles of books such as *Assessing teacher effectiveness* (Campbell, Kyriakides, Muijs and Robinson, 2004) and, in ELT, *Evaluating teacher effectiveness in ESL/EFL contexts* (Coombe et al., 2007). Teacher effectiveness refers to the impact of teaching on students, either with specific reference to learning outcomes or more generally, as this definition illustrates:

> *the collection of characteristics, competencies, and behaviors of teachers at all educational levels that enable students to reach desired outcomes, which may include the attainment of specific learning objectives as well as broader goals such as being able to solve problems, think critically, work collaboratively, and become effective citizens. (Hunt, 2009, p. 1)*

In a similar vein, Campbell, Kyriakides, Muijs and Robinson (2004, p. 4) say that teacher effectiveness is 'the power to realise socially valued objectives agreed for teachers' work, especially, but not exclusively, the work concerned with enabling students to learn'.

Another distinction that arises in the literature on teacher evaluation is that between **teacher quality** and **teaching quality**. Darling-Hammond (2013) argues that the former is what the teacher is capable of doing and the latter is what they are able to do in a given context (the idea being that a good-quality teacher may, in adverse circumstances, not demonstrate good-quality teaching). The same distinction

is reflected in the terms **teacher performance** and **teacher competence**. 'Performance' refers specifically to teachers' observed instructional behaviours in the classroom (see, for example, Miao, Reynolds, Harris and Jones, 2015; Steinberg and Garrett, 2015). 'Competence' 'is the extent to which the teacher possesses the knowledge and skills (competencies) defined as necessary or desirable qualifications to teach' (Dunkin, 1997).

Teacher evaluation, then, is the process through which judgements about the quality of teachers are made. Assessing teachers' performance in the classroom and its impact on students are core elements in this process. However, as suggested in the definitions above, in evaluating teachers it is also important to consider broader elements of teacher quality such as planning, reflection, professional development and contributions to school effectiveness more generally (see Section 3.3 for further discussion of teacher evaluation criteria).

This review focuses on different ways of evaluating teachers. However, because decisions about how to evaluate teachers need to be informed by and aligned with other components in a teacher evaluation framework, the next section will introduce some of these broader components.

---

[2] It must be noted here that teacher assessment also refers to the process through which students are assessed by their teachers (Johnson, 2013). This sense of the term is not relevant to this review.

# 3

# Teacher evaluation frameworks

Various frameworks for teacher evaluation systems are available in the literature. That presented by Isoré (2009) (see Figure 1) suggests that a coherent approach to teacher evaluation requires decisions about the following issues: agencies and stakeholders involved, the scope of the evaluation (for example, which teachers to evaluate), evaluators (who they will be), criteria and standards for evaluation, methods and instruments, and uses of teacher evaluation (formative and summative).
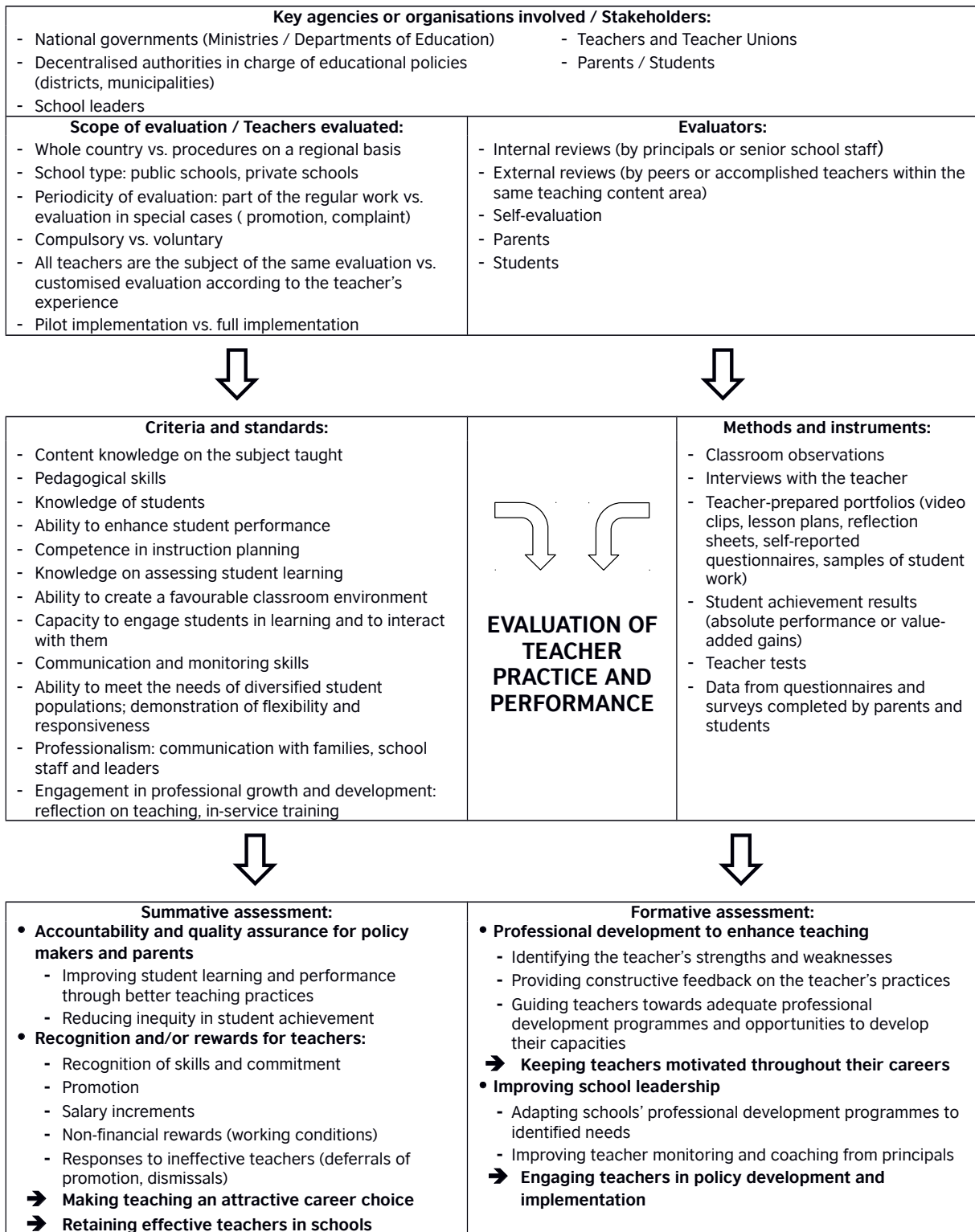
---

**Key agencies or organisations involved / Stakeholders:**

- National governments (Ministries / Departments of Education)
- Decentralised authorities in charge of educational policies (districts, municipalities)
- School leaders
- Teachers and Teacher Unions
- Parents / Students

---

| Scope of evaluation / Teachers evaluated: | Evaluators: |
|---|---|
| - Whole country vs. procedures on a regional basis<br>- School type: public schools, private schools<br>- Periodicity of evaluation: part of the regular work vs. evaluation in special cases ( promotion, complaint)<br>- Compulsory vs. voluntary<br>- All teachers are the subject of the same evaluation vs. customised evaluation according to the teacher's experience<br>- Pilot implementation vs. full implementation | - Internal reviews (by principals or senior school staff)<br>- External reviews (by peers or accomplished teachers within the same teaching content area)<br>- Self-evaluation<br>- Parents<br>- Students |

⇩       ⇩

| Criteria and standards: | EVALUATION OF TEACHER PRACTICE AND PERFORMANCE | Methods and instruments: |
|---|---|---|
| - Content knowledge on the subject taught<br>- Pedagogical skills<br>- Knowledge of students<br>- Ability to enhance student performance<br>- Competence in instruction planning<br>- Knowledge on assessing student learning<br>- Ability to create a favourable classroom environment<br>- Capacity to engage students in learning and to interact with them<br>- Communication and monitoring skills<br>- Ability to meet the needs of diversified student populations; demonstration of flexibility and responsiveness<br>- Professionalism: communication with families, school staff and leaders<br>- Engagement in professional growth and development: reflection on teaching, in-service training | | - Classroom observations<br>- Interviews with the teacher<br>- Teacher-prepared portfolios (video clips, lesson plans, reflection sheets, self-reported questionnaires, samples of student work)<br>- Student achievement results (absolute performance or value-added gains)<br>- Teacher tests<br>- Data from questionnaires and surveys completed by parents and students |

⇩       ⇩

| Summative assessment: | Formative assessment: |
|---|---|
| • **Accountability and quality assurance for policy makers and parents**<br>  - Improving student learning and performance through better teaching practices<br>  - Reducing inequity in student achievement<br>• **Recognition and/or rewards for teachers:**<br>  - Recognition of skills and commitment<br>  - Promotion<br>  - Salary increments<br>  - Non-financial rewards (working conditions)<br>  - Responses to ineffective teachers (deferrals of promotion, dismissals)<br>➔ **Making teaching an attractive career choice**<br>➔ **Retaining effective teachers in schools** | • **Professional development to enhance teaching**<br>  - Identifying the teacher's strengths and weaknesses<br>  - Providing constructive feedback on the teacher's practices<br>  - Guiding teachers towards adequate professional development programmes and opportunities to develop their capacities<br>➔ **Keeping teachers motivated throughout their careers**<br>• **Improving school leadership**<br>  - Adapting schools' professional development programmes to identified needs<br>  - Improving teacher monitoring and coaching from principals<br>➔ **Engaging teachers in policy development and implementation** |

**Figure 1:** Conceptual framework for teacher evaluation (Isoré, 2009, p. 32)

Similarly, Santiago and Benavides (2009) present a framework (see Figure 2) with six components which can be framed as questions:
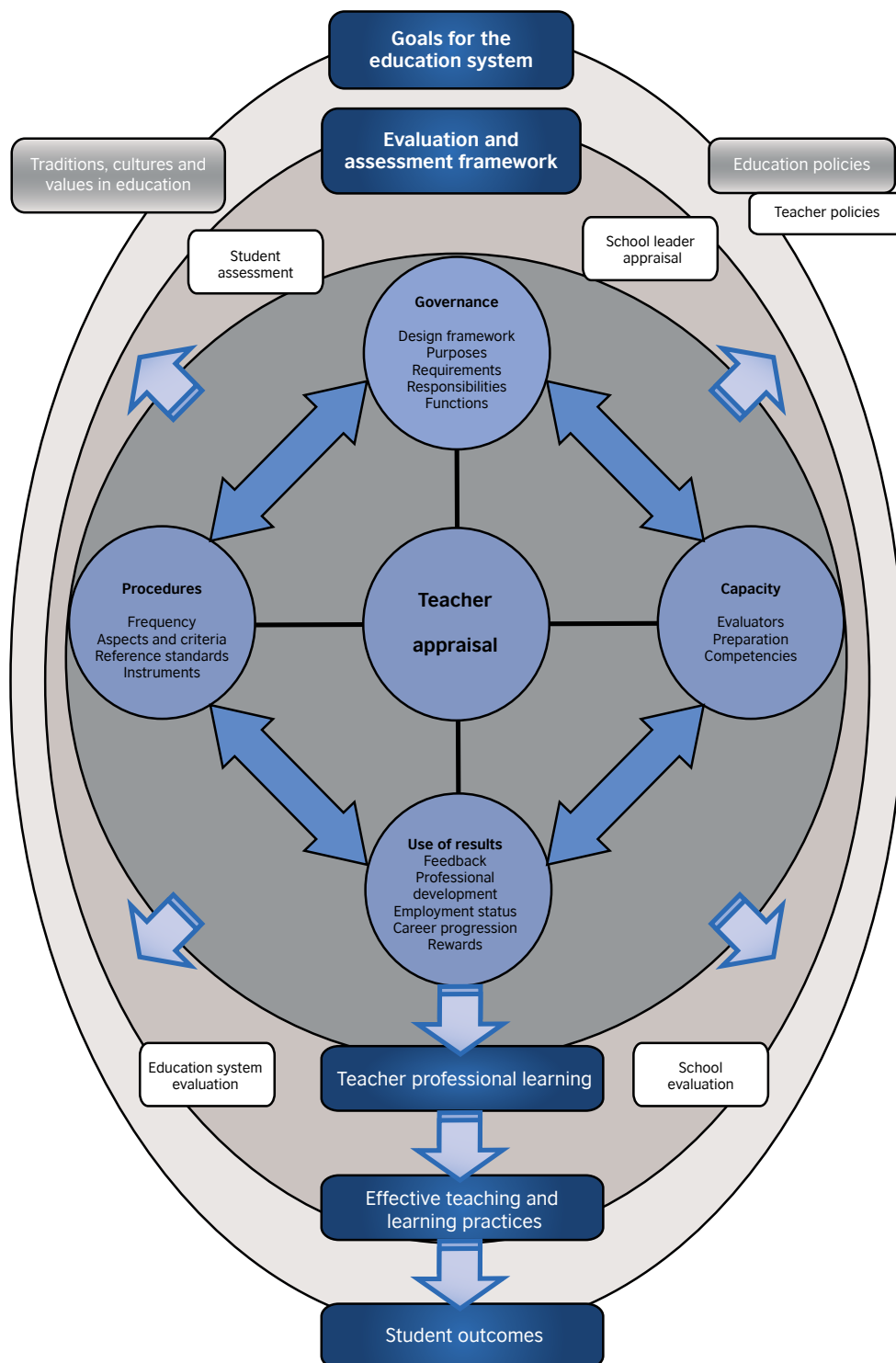
- Who will conduct the evaluation and use the results?

- Who will be assessed?

- What aspects of teaching will be evaluated?

- How will the evaluation take place?

- What will the purposes of the evaluation be?

- Who will be involved in the implementation of the evaluation?
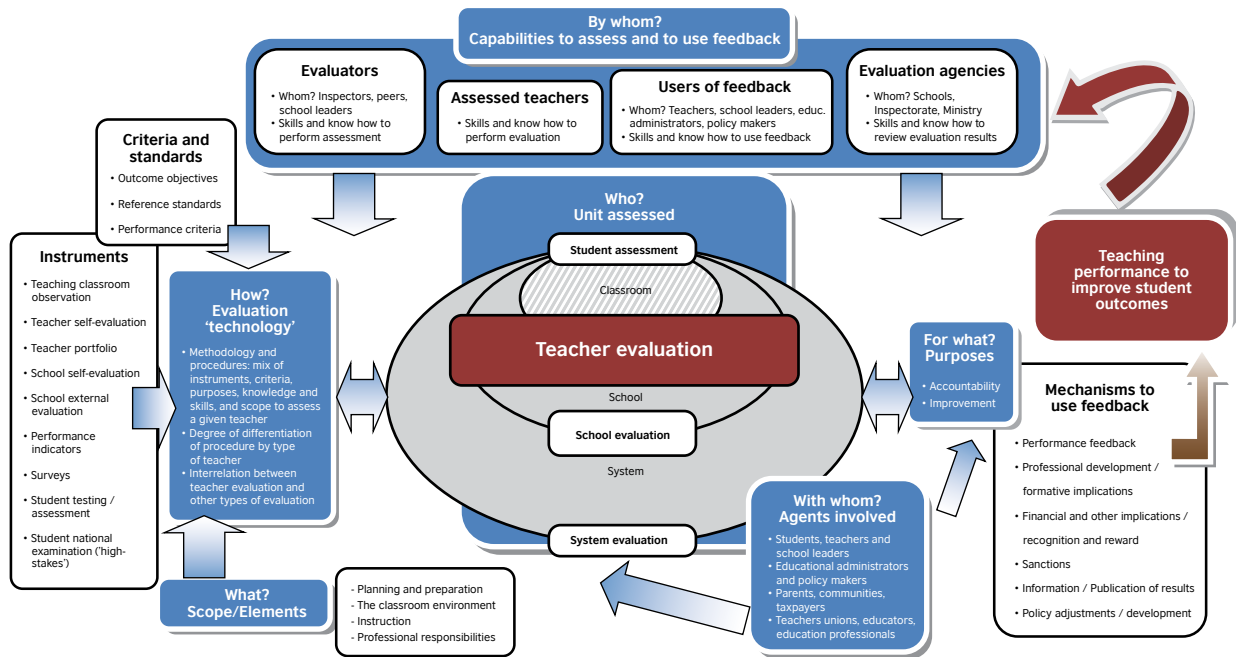


*OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes*

**Figure 2:** Teacher evaluation framework (Santiago and Benavides, 2009, p. 5)

A more recent example of a teacher evaluation framework is being used in the OECD's current programme of reviews of evaluation and assessment (OECD, 2015). This framework (see Figure 3) discusses teacher evaluation in relation to four broad considerations: governance (for example, design and responsibilities), capacity (such as preparation of evaluators), procedures (such as who and what to assess) and use of results (for example, professional development or decisions about salary or employment).



Goals for the education system

Evaluation and assessment framework

Traditions, cultures and values in education

Education policies

Teacher policies

Student assessment

School leader appraisal

**Governance**
Design framework
Purposes
Requirements
Responsibilities
Functions

**Procedures**
Frequency
Aspects and criteria
Reference standards
Instruments

**Teacher appraisal**

**Capacity**
Evaluators
Preparation
Competencies

**Use of results**
Feedback
Professional development
Employment status
Career progression
Rewards

Education system evaluation

Teacher professional learning

School evaluation

Effective teaching and learning practices

Student outcomes

SYNERGIES FOR BETTER LEARNING: AN INTERNATIONAL PERSPECTIVE ON EVALUATION AND ASSESSMENT © OECD 2013

**Figure 3:** OECD Teacher evaluation framework (OECD, 2013a, p. 275)

In order to provide a framework for the discussion of teacher evaluation strategies in Section 4, specific components of teacher evaluation systems will now be introduced below.

## 3.1 Purpose of teacher evaluation

Teacher evaluation can fulfil a range of purposes and a key distinction is that between accountability and development.[3] In the former, teacher evaluation is used summatively (for example, to make decisions about status or salary) while, in the latter, the evaluation process helps teachers improve (i.e. teachers receive feedback on their performance, and support in helping them address areas for development). **Contemporary views of teacher evaluation repeatedly stress the importance of a formative or developmental perspective.** For example, Darling-Hammond (2013, p. 153) notes that 'evaluation should be accompanied by useful feedback, and connected to professional development opportunities that are relevant to teachers' goals and needs' while Marzano and Toth (2013, p. 14) note that 'an effective evaluation system should help teachers teach better'. Similarly, OECD analyses (for example, OECD, 2013a) highlight the importance of ensuring that teacher evaluation informs professional development and cites Korea and Singapore as contexts where this takes place. In the UK, the Department for Education's teacher appraisal model (Department for Education, 2012) also stresses the importance of supporting teachers (though it does include, too, procedures for dealing with underperforming teachers). Teachers do, in fact, seem to value teacher evaluation that has a formative function; citing Howard and McCloskey (2001), King (2015) notes that, especially among experienced teachers, one major aversion to evaluation is that it does not promote professional growth, while Marzano and Toth (2013) cite a survey in which 76 per cent of some 3,000 teachers said that, while measurement and development are both important in teacher evaluation, development should predominate.

**The importance of formative teacher evaluation, though, should not detract from its equally important summative function.** As noted by Isoré (2009, p. 7), 'In its summative form, evaluation firstly responds to the needs of assuring that teaching is directed towards student achievement. It also provides opportunities for social recognition of teachers' skills and commitment to work. These are two major concerns in our knowledge societies'. Summative teacher evaluation holds teachers accountable for their performance and can have consequences such as career advancement, bonus pay, or sanctions for underperformance (OECD, 2013b).

The formative and summative purposes of teacher evaluation are often seen to be contradictory and it has been claimed that given such tensions it is the summative purposes of teacher evaluation that often predominate (Campbell et al., 2004). Evidence from teacher evaluation in South Africa supports this point; the Integrated Quality Management System (IQMS) provides scope for formative and summative teacher evaluation, but its implementation has focused largely on performance measurement for pay and promotion (The Centre for Development and Enterprise, 2015). (See also Avalos and Assael, 2006 for a discussion of formative and summative teacher evaluation in Chile.) Santiago and Benavides (2009) also discuss the challenges of combining formative and summative teacher evaluation and provide advice on the kinds of conditions

that are likely to support each (see Box 1). One potential drawback of teacher evaluation which is summative is that it can discourage teacher collaboration, which is widely recognised as a feature of effective schools (for examples from Ontario, Canada, see Lieberman, Campbell and Yashkina, 2017).

---

**Box 1:** Formative and summative teacher evaluation (Santiago and Benavides, 2009, pp. 8–9)

Teacher evaluation for improvement purposes is likely to benefit from conditions such as:

■ A non-threatening evaluation context

■ A culture of mutually providing and receiving feedback

■ Clear individual and collective objectives with regard to improving teaching within the school as well as a sharing of school objectives

■ Simple evaluation instruments such as self-evaluation forms, classroom observation, and structured interviews

■ A supportive school leadership

■ Opportunities to enhance competencies, as well as resources and means to improve practice

■ Teacher evaluation integrated in a system of school self-evaluation and quality assurance.


Teacher evaluation for accountability is likely to benefit from conditions such as:

■ An independent and objective assessment of the teacher's performance

■ National-level standards and criteria across schools

■ An evaluation component external to the school and more formal processes

■ Well-established rules regarding the consequences of the evaluation

■ Clear individual objectives with regard to all aspects of a teacher's performance

■ Well-trained, competent evaluators of teaching performance

■ Impact on professional development plan

■ Possibilities for appeal for teachers who feel they have not been treated fairly.

---

[3] While teacher evaluation can most obviously support the growth of individual teachers, it can also contribute to the development of schools and educational systems more generally.

## 3.2 Participants, frequency and obligation

Individual teachers are the focus of teacher evaluation, but beyond that obvious point various decisions need to be made about who will be evaluated, how often, and about whether the process will be compulsory or voluntary.

*Participants.* Will all teachers in an institution, organisation or system be evaluated or only those meeting certain criteria? Will distinctions be made according to, for example, how experienced teachers are? Abdelaziz et al. (2016), in their case studies of teacher evaluation in different countries, note that 'many of the countries share a philosophy that long-standing teachers, with or without the formal status of tenure, should be given more freedom of teaching, with much less evaluation, supervision or corrective steps required, unless infringements are deemed significant' (p. 6). Such an approach erroneously equates experience with expertise (see Goodwyn, 2017 for a recent discussion of expert teachers). It also assumes that teacher evaluation is only a corrective activity rather than a process that seeks to raise the quality of all teachers.

*Frequency.* How often will teachers be evaluated? Where national frameworks exist, most OECD countries report that teachers should be evaluated at regular intervals, most typically every year (OECD, 2014a). For example, the Australian Teacher Performance and Development Framework requires an annual appraisal process for all teachers.[4] The word 'annual', though, should not imply that teacher evaluation is a one-off event and it should be seen as an on-going process, with evidence of teachers' work being collected at different times during the year. Zhang and Ng (2017, p. 214) illustrate how this works in Shanghai, noting that while teacher evaluation:

> occurs at the end of each year, and mainly seeks to generate a summative evaluation of individual teachers' morality,[5] competence, achievements, and diligence, key appraisal activities (such as lesson observation, student evaluation, analysing student examination results, and checking teachers' tasks) are conducted throughout the whole year to collect data regarding teachers.

Globally, it is likely that in practice the frequency of teacher evaluation varies significantly; for example, according to OECD (2014b), over 60 per cent of teachers in Iceland have never been observed for appraisal purposes; in contrast, Liu and Zhao (2013) describe a school in China where teacher evaluation was conducted every month.

*Obligation.* Will teacher evaluation be voluntary or compulsory? The answer to this question will vary depending on the purposes of teacher evaluation. According to OECD (2015, p. 493), 'while regular appraisal, appraisal for the completion of probation, and appraisal for teacher registration is mandatory in the majority of [OECD] countries … appraisal for promotion and reward schemes are usually voluntary'. Chile's National System of School Performance Evaluation illustrates this: while regular appraisal is compulsory for all teachers, additional teacher evaluation pathways that offer salary rewards are also available on a voluntary basis (Santiago, Benavides, Danielson, Goe and Nusche, 2013).

## 3.3 Evaluation criteria

Another critical component of teacher evaluation systems is that they will specify the criteria against which assessments are made. Effective teacher evaluation is not possible unless the criteria for evaluation have been clearly defined. These criteria can take many forms but commonly appear as competency frameworks or professional standards, which can vary in their prescriptiveness and intended use but which always reflect particular beliefs about what teacher competence entails.

Few examples of specific ELT frameworks exist within state education systems and it is more usual for a common set of standards or competences to be deployed for all teachers irrespective of subject. Exceptions to this trend do exist in, for example, Malaysia and Vietnam, where the Ministries of Education have developed separate frameworks for teachers of English. A range of additional 'global' frameworks also exist (see Box 2) which were all designed to support teacher development and thus have a clear formative purpose (see Section 3.1). For example,

---

**Box 2:** Frameworks and standards for foreign language teaching

**British Council Continuing Professional Development Framework**

Describes 12 professional practices for teachers of English and four stages of development (Awareness, Understanding, Engagement and Integration).

Source: www.teachingenglish.org.uk/article/british-council-cpd-framework

**Cambridge Assessment English Teaching Framework**

Organised around five categories of teaching knowledge and skills which can be assessed at four levels (Foundation, Developing, Proficient, Expert).

Source: www.cambridgeenglish.org/teaching-english/cambridge-english-teaching-framework

**European Profiling Grid**

Covers four categories of language teachers' professional practice: Training and Qualifications, Key Teaching Competences, Enabling Competences and Professionalism. These are assessed against six stages of professional development.

Source: www.epg-project.eu/the-epg-project

---

[4] www.aitsl.edu.au/tools-resources/resource/australian-teacher-performance-and-development-framework
[5] While the literature on teacher quality does not ignore the ethical dimension of teaching, the explicit reference to morality here reflects the prominent attention this aspect of teachers' work receives in contexts such as China.

the designers of the European Profiling Grid (EPG) note 'it must be stressed that the EPG does not set out to be a set of standards or rules to be imposed on language teachers … the EPG should not be used as an instrument to direct, impose, restrict, harmonise, reward or penalise teachers' (Mateva, Vitanova and Tashevska, 2011, p. 4). These frameworks also lend themselves to teacher self-evaluation (see Section 4.5).

Further frameworks are available that have been designed for use in pre-service language teacher education contexts, such as the TESOL/CAEP (formerly NCATE) P–12 Teacher Education Program standards and Standards for Short-Term TEFL/TESL Certificate Programs (for a discussion of some of these, see Nunan, 2007), the European Portfolio for Student Teachers of Languages (EPOSTL) and the European Profile for Language Teacher Education. Despite their focus on pre-service teacher education, the standards they list can also inform the evaluation of practising teachers. A book on developing EFL professional teaching standards has also been published by the TESOL International Association (Kuhlman and Knežević, undated). The standards for foreign language learning established by the American Council on the Teaching of Foreign Languages (ACTFL) have been used in the USA to define standards for pre-service foreign language teaching.[6]

One key aspect of teacher competence in ELT is English language proficiency and this is therefore a legitimate focus for teacher evaluation. It is recognised that in several ELT contexts around the world, particularly in primary schools, teachers of English have modest levels of English and this is a significant barrier to high-quality English language teaching. Decisions about evaluating teachers' English must be underpinned by an understanding of the kind of English teachers need (see, for example, Freeman, Katz, Garcia Gomez and Burns, 2015) and of the level of English teachers require (Hayes, 2014, for example, recommends B2 to C1 on the Common European Framework of Reference).[7] The testing of teachers' English is discussed further in Section 4.6

Empirical accounts of how ELT frameworks have been validated and used in evaluation systems for practising state school teachers remain limited. The TESOL International Association's professional standards have underpinned work in a number of countries, especially in pre-service contexts (for an example from Egypt, see McCloskey, Thornton and Touba, 2007). Murphey and Yaode (2007), though, do describe a project in which these standards also informed those for practising teachers in China and which were organised around eight domains:

- Knowing students

- Appreciating attitudes

- Planning, delivering and reflecting on instruction

- Constructing knowledge of languages, language learning and critical thinking

- Exploring and applying culture

- Assessing teaching and learning

- Connecting beyond the classroom

- Expanding professional horizons.

In education generally, there is also an extensive literature focused on professional standards and competency frameworks. For example, in the USA, the National Board for Professional Teaching Standards (NBPTS) has produced a set of influential standards (National Board for Professional Teaching Standards, 2016) which are based on five propositions about what teachers should know and be able to do:[8]

- Proposition 1: Teachers are committed to students and their learning

- Proposition 2: Teachers know the subjects they teach and how to teach those subjects to students

- Proposition 3: Teachers are responsible for managing and monitoring student learning

- Proposition 4: Teachers think systematically about their practice and learn from experience

- Proposition 5: Teachers are members of learning communities.

To take another example from outside ELT, the Interstate Teacher Assessment and Support Consortium (InTASC), again in the USA, also developed a set of teaching standards which are grounded in contemporary understandings of effective teaching and learning (Council of Chief State School Officers, 2013). Ten standards are identified (see Box 3 for the domains covered) and for each expected teacher performance, knowledge and critical dispositions are defined.

---

**Box 3:** Domains in InTASC teaching standards

1  Learner development

2  Learning differences

3  Learning environments

4  Content knowledge

5  Application of content

6  Assessment

7  Planning for instruction

8  Instructional strategies

9  Professional learning and ethical practice

10  Leadership and collaboration

---

[6] www.actfl.org/sites/default/files/publications/standards/World-ReadinessStandardsforLearningLanguages.pdf
[7] www.coe.int/en/web/common-european-framework-reference-languages/
[8] See National Board for Professional Teaching Standards (2010) for professional standards relevant to the teaching of English as a new language.

In Asia, too, there has been growing attention to competency frameworks and Box 4 presents an example developed with input from stakeholders in eleven countries in South East Asia.

---

**Box 4:** Competency framework for South East Asian teachers of the twenty-first century (Seameo Innotech, 2010, pp. 92–93)

**A**  Facilitating the development of learners' life and career skills

**B**  Facilitating learning

**C**  Preparing appropriate lesson plans in line with the school vision and mission

**D**  Creating a conducive learning environment

**E**  Developing and utilizing teaching and learning resources

**F**  Developing higher order thinking skills (HOTS)

**G**  Enhancing ethical and moral values

**H**  Assessing and evaluating learner performance

**I**  Engaging in professional development

**J**  Networking with stakeholders especially with parents

**K**  Managing students' welfare and other tasks

---

The source of this framework also notes that the 11 countries surveyed were at different stages in the process of developing, implementing and evaluating their own national frameworks, with little evidence of how such frameworks were operating in practice. This is an important point about teacher competency frameworks and teacher education systems more generally; while establishing them is important, it is how they are enacted which matters most and this is an issue where much more research (including through case studies in specific contexts) is required.

In terms of the specific aspects of teaching that teacher evaluation assesses, planning and preparation, the classroom environment and instruction are very common (OECD, 2013b). However, as the InTASC and NBPTS standards indicate and, as noted in Section 2 above, **assessments of teacher quality should not be limited to instructional matters and should also encompass areas of teachers' work such as professional development and collaboration**. This broader conception of teacher quality is reflected in contemporary discussions of teacher evaluation; for example, Darling-Hammond (2013) gives these examples of the kinds of professional contributions that can be included in teacher evaluation schemes:

- Developing and sharing curricula

- Supporting colleagues through peer observation

- Mentoring and coaching

- Leadership roles in school initiatives

- Outreach to parents

- Sharing instructional practices.

Teaching standards may differ according to the teacher being evaluated. In New Zealand, for example, there are different standards for beginner teachers, classroom teachers and experienced teachers (Nusche, Laveault, MacBeath and Santiago, 2012). In Ontario, Canada, the Teacher Performance Appraisal System also distinguishes between new teachers and experienced teachers.

Contemporary teaching frameworks (such as that by InTASC above) are evidence-based in the sense that they focus on professional behaviours and strategies which have been shown by teacher effectiveness research to impact on student learning. An exhaustive discussion of these and other standards is beyond the scope of this review; the key point here, though, is that **effective teacher evaluation must be based on clearly defined criteria which are informed by explicit statements about what teachers are expected to know and do**.

One further recurrent point about teaching standards made in the literature is that teacher involvement in their development is highly desirable:

> *For the teaching standards to be relevant and owned by the profession, it is essential that the teaching profession takes a lead role in developing and taking responsibility for them. The participation of teachers in designing standards (and procedures) for teacher appraisal is essential to the effectiveness of any appraisal system. (OECD, 2013a, p. 297)*

In both Australia and New Zealand, for example, teaching standards were developed through consultation with the teachers and employers. However, it is important that consultation is a meaningful exercise to which teachers are able to contribute fully and where their opinions carry weight. For example, Purdon (2003) notes that in Scotland the introduction of a new framework for continuing professional development was characterised by numerous consultation exercises; however, it is claimed that their quality and the value attached to teachers' responses were insufficient and that control of the process remained largely with the authorities.

Box 5 summarises key points that have been made so far about teacher education frameworks and their key components. As explained earlier, this broader preliminary discussion was necessary because decisions about how to evaluate teachers (which the rest of the report will now focus on) need to be informed by and aligned with other components in a teacher evaluation system. Further discussions of effective teacher evaluation systems are provided in Darling-Hammond (2013), Marzano and Toth (2013), Danielson and McGreal (2000), Bruns and Luque (2015), Glazerman et al. (2011), Kaufman (2007), Stoynoff (2007), Murdoch (2000), Hayes, Chang and Imm (2011) and Quirke (2015). Various OECD reports (OECD, 2013a, 2013b) are an additional source of guidance on the design and

implementation of teacher evaluation systems in a range of countries.

---

**Box 5:** Overview of teacher evaluation

1    Teacher evaluation is the process of assessing teacher quality.

2    Teacher quality includes performance in the classroom (teaching quality) as well as broader aspects of teachers' professional work.

3    Effective teacher evaluation involves the coherent specification of purposes, processes, stakeholders, and use of results.

4    Teacher evaluation can serve formative and summative purposes and these may call for different approaches to the assessment of teacher quality.

5    Effective teacher evaluation will be based on explicit criteria about what teachers should know and be able to do at particular points in their career.

6    Teacher evaluation systems can be enhanced by stakeholder involvement in their development and implementation.

7    An effective teacher evaluation system discriminates between more and less effective teachers.

8    Teacher evaluation should contribute to more effective teaching and learning.

---

In many ELT contexts teacher evaluation has been traditionally associated with the process of teacher or school inspection.[9] From the discussion so far (and from what follows in the rest of this review) it should be clear that effective teacher evaluation extends far beyond the kinds of one-off, summative and very often stressful events (Hopkins et al., 2016) that inspection is associated with. (In response to such concerns, as Stevens (2016) notes, school inspection in the UK no longer involves the evaluation of individual teachers.) Without denying a role for summative assessments of a more formal kind, this review advances a view of teacher evaluation as a process that is positive and constructive, involves multiple evaluators, and draws on several sources of rigorous evidence collected over time.

---

[9]  For an example from France, see www.sici-inspectorates.eu/getattachment/9e11ce92-9c36-4e1e-af1e-a43d305437d7

# 4

# Teacher evaluation strategies

An important principle for teacher evaluation that recurs in the contemporary literature (such as the collection of papers in Grissom and Youngs, 2016) is that it should utilise multiple sources of evidence. In this section, different ways of collecting evidence of teacher quality will be discussed. As Box 6 shows, a wide range of options are available which imply not just different instruments but different evaluators, such as head teachers, administrators, students, parents, and teachers themselves (in both self- and peer evaluation).

---

**Box 6:** Strategies for teacher evaluation

**1**   Classroom observation

**2**   Student outcomes

**3**   Teaching portfolios

**4**   Student evaluations of teaching

**5**   Self-evaluation

**6**   Teacher tests

**7**   Professional conversations

**8**   Peer evaluation

**9**   Parent feedback

---

Multiple measures are recommended in teacher evaluation for a variety of reasons:

- to adequately capture the complexity of teaching (in a way that no single measure can)

- to give teachers different opportunities to show what they know and can do

- to minimise the bias and lack of reliability that any one measure may have.

No teacher evaluation system, of course, will employ all of the measures available, and, as Isoré (2009, p. 20) notes, 'while the multiplication of instruments and evaluators is more likely to provide a solid basis to evaluate teachers, limited resources make trade-offs inevitable'. This is even more likely when teacher evaluation takes place at scale given that certain strategies, such as teaching portfolios (see Section 4.3), are more time-consuming to administer than those which rely on quantitative data. It is also important that evidence from multiple sources 'is collected over multiple points in time' (Marzano and Toth, 2013, p. 13). This emphasises a point made earlier about **the need to see teacher evaluation as an extended process and not an event that happens at a single point in time**.

One research project in the USA that emphasised the use of different sources of evidence in assessing teacher effectiveness was the large-scale Measures of Effective Teaching (MET) project (Kane, Kerr and Pianta, 2014), where teachers were assessed using observations, student ratings and student test scores. This project concluded that 'a generally balanced set of different measures was seen to produce more stable results and a better indication of student learning on a range of assessments than one that gives a preponderance of weight to a single measure' (Archer, Kerr and Pianta, 2014, p. 3).

On the following page, Box 7 describes some teacher evaluation systems that utilise multiple measures. In all cases, these apply to teachers across subjects and are not specific to ELT.

**Box 7:** Multiple measures in teacher evaluation

**Chile**
These are the four instruments and their relative weights in the overall assessment score:

- Self-assessment: structured questionnaire to prompt teachers to reflect on their teaching performance (10 per cent)

- Portfolio: teachers submit a written lesson plan and a professionally made videotape of one of their classes (60 per cent)

- Peer interview: structured questionnaire to examine how the teacher being evaluated would handle different pedagogical challenges, applied by a similar teacher trained to conduct these interviews following a set protocol (20 per cent)

- Third-party reports: structured questionnaire completed by the school director and a district pedagogical supervisor (10 per cent). (Bruns and Luque, 2015, pp. 186–187)

**Ontario, Canada**
Although, depending on whether the teacher is 'new' or 'experienced', some components are applied differently … the following components are common to the appraisal of both new and experienced teachers: Classroom observation of the teacher by the appraising principal.

Appraisal meetings that promote professional dialogue between the principal and the teacher. A principal must arrange a pre-observation meeting with the teacher to prepare for the classroom observation and a post-observation meeting to discuss what went on during the observation. The meetings provide opportunities for reflection and collaboration to promote growth and improvement. (Ministry of Education, 2010, p. 9)

**Korea**
The key features of the Teacher Appraisal for Professional Development are summarised as follows: a multi-dimensional appraisal method is adopted which involves the input of the whole education community. Principals, vice principals and teachers function as evaluators, while students and parents provide information on satisfaction levels by filling in questionnaires composed of checklist questions and descriptive opinion writing. (Kim et al., 2010, p. 54)

**Kosovo**
A new teacher evaluation scheme uses the following measures:

- Teacher's self-assessment – 10 per cent/12 points

- Assessment from school director – 30 per cent/36 points

- External assessment/class observation – 30 per cent/36 points

- Portfolio, i.e. evidence of planning and implementation of teaching and learning as a whole – 30 per cent /36 points.

(www.ec.europa.eu/dgs/education_culture/repository/education/events/2016/wbp/teacher-evaluation-tools-criteria_en.pdf)

**Shanghai**
Teacher evaluation is based on:

- Lesson observation

- Student evaluations

- Analysis of examination results

- Inspection of teachers' work.

(Zhang and Ng, 2017)

## 4.1 Classroom observation

Classroom observation is a very widely used measure of teacher quality. It provides direct evidence of what teachers do and is thus a key component of performance-based teacher evaluation systems (for an analysis of such systems see Shakman et al., 2012). Observations can serve formative and summative purposes and, depending on the level of formality involved, may be conducted by school principals, external observers such as inspectors, or peers (see 4.7.2 for peer evaluation).

Observation, as conventionally used for teacher evaluation, has, though, been criticised on a number of grounds:

■ Teachers find it threatening and intrusive (King, 2015).

■ Observer judgements can be influenced by their prior knowledge of and relationship with the teacher (Bell et al., 2014).

■ Observation tools may be poorly designed (OECD, 2013a).

■ Observers (often school principals) may lack the capacity and training to judge teachers appropriately (Darling-Hammond, 2013).

■ One-off observations provide a very partial and unrepresentative picture of what teachers can do (Campbell et al., 2004).

■ Observations tend to focus on accountability rather than professional development (Marzano and Toth, 2013).

■ Observation reduces teaching to a superficial set of skills and behaviours (O'Leary, 2016a).

In response to such concerns, **improving the value of classroom observation – especially through improved observation tools and evaluator training – has become a major focus in the teacher evaluation literature**. Below I review some key contemporary perspectives on the use of observation in teacher evaluation.

### 4.1.1 Observation tools

According to Pianta and Hamre (2016), observation tools should be standardised, reliable and valid. Standardised tools ensure that observation procedures, including how scores are assigned, are uniformly followed. Reliability (see 4.1.2 below) refers to the consistency of measurement while validity refers to the extent to which a tool measures what it claims to measure. One way to assess the validity of observational measures is to examine their associations with other assessments of teacher quality, such as student achievement (see Section 4.2) or student evaluations of teaching (see Section 4.4).

Box 8 lists some observation tools that are widely used and seen to meet the criteria discussed above. In the Measures of Effective Teaching Project (see, for example, Mihaly and McCaffrey, 2014), five observation tools that are widely used in the USA were employed; two of these (CLASS and Framework for Teaching) were generic while three were subject-specific: Protocol for Language Arts Teaching Observations (PLATO), Mathematical Quality of Instruction (MQI) and Quality Science Teaching (QST). Marzano and

---

**Box 8:** Observation tools

**Classroom Assessment Scoring System (CLASS)**
The Classroom Assessment Scoring System is a tool for analysing the quality of teacher–student interactions in the classroom. It produces qualitative ratings of teacher performance on a scale of 1–7 across three broad domains: emotional support, classroom organization, and instructional support.

www.worldbank.org/en/programs/sief-trust-fund/brief/the-classroom-assessment-scoring-system-class

**Framework for Teaching (FFT)**

The Framework for Teaching (FFT) is a research-based protocol that divides the complex activity of teaching into 22 components (and 76 smaller elements) clustered into four domains of teaching responsibility: Planning and preparation (Domain 1), Classroom environment (Domain 2), Instruction (Domain 3), and Professional responsibilities (Domain 4).

www.k12education.gatesfoundation.org/resource/danielsons-framework-for-teaching-for-classroom-observations/

**The Protocol for Language Arts Teaching Observations (PLATO)**

The Protocol for Language Arts Teaching Observations (PLATO) is a classroom observation protocol designed to capture features of English/Language Arts (ELA) instruction. The PLATO protocol covers four instructional domains – Disciplinary demand of classroom talk and activity, Contextualizing and representing content, Instructional scaffolding, and Classroom environment – and 13 elements of instruction identified by research on adolescent literacy and effective instruction in ELA.

www.platorubric.stanford.edu/

**Stallings Classroom Snapshot**

The Stallings classroom observation system, also known as the Stallings Classroom Snapshot, is a questionnaire and protocol for timed observations that produce quantitative data about interactions of teachers and students in classrooms. It collects information about:

■ Teachers' use of time: for instruction; classroom management; or off-task (out of the classroom or in the classroom in social interaction)

■ Teachers' use of different learning activities: reading aloud; demonstration/lecture; discussion/question and answer; practice and drill; assignment/class work; copying

■ Teachers' ability to keep students engaged.

www.worldbank.org/en/programs/sief-trust-fund/brief/the-stallings-classroom-snapshot

Toth (2013, pp. 43–44) present an observation tool that consists of 41 specific elements which they claim is comprehensive ('it includes a wide variety of instructional strategies that are associated with student achievement') and specific ('the model identifies classroom strategies and behaviours at a very granular level').

Many of these instruments are quite detailed and require extensive training for users; this has implications for the cost of teacher evaluation and questions have thus been raised about whether the quality of teaching can be assessed as effectively but less expensively and more rapidly using instruments with fewer items. Gargani and Strong (2014) claim this is possible using their Rapid Assessment of Teaching Effectiveness (RATE) tool, which includes just six items, requires only four hours of training and can be applied to 20-minute lesson observations. Their claims have not passed uncontested (Good and Lavigne, 2015) and debates of this kind are further evidence of the high levels of current interest in teacher evaluation generally and classroom observation more specifically.

One feature that these systematic observation instruments share is that they use a rating scale which allows for a lesson, or specific aspects of it, to be graded. Although many options are available, scales cover a continuum of performance from less to more satisfactory; for example, the Quality Science Teaching (QST) tool uses a four-point scale: little or no evidence; limited evidence; clear evidence; consistently strong evidence (Schultz and Pecheone, 2014), while Marzano and Toth's instrument described above uses a five-point scale for evaluating the quality of specific aspects of teaching: innovating – applying – developing – beginning – not using. The Framework for Teaching uses this four-point scale: unsatisfactory, basic, proficient and distinguished, while, to take one final example, the observation tool developed with input from several countries as part of the International Comparative Analysis of Learning and Teaching project (de Jager, Coetzee, Maulana, Helms-Lorenz and van de Grift, 2017; van de Grift, 2007, 2014; van de Grift, Chun, Maulana, Lee and Helms-Lorenz, 2017) uses this scale for evaluating specific elements of observed teaching: (1) predominantly weak; (2) more weaknesses than strengths; (3) more strengths than weaknesses; and (4) predominantly strong.

### 4.1.2 Observer reliability

As noted above, observer reliability is a key element in the effective use of classroom observation for teacher evaluation. Problems are, though, noted in the literature (see Bell et al., 2014) with variations in the way different observers rate the same lesson (inter-rater reliability) or in how an observer uses the same set of criteria on different occasions (intra-rater reliability). The Measures of Effective Teaching project, in which thousands of classroom observations were conducted and analysed, concluded that the reliability of classroom observations could be improved through rigorous training of observers and using different observers to assess multiple lessons by the same teacher (Ho and Kane, 2013). Standardisation is also essential as it provides the basis for observer training. Pianta and Hamre (2016) identify three areas of standardisation to consider: the training protocol,

parameters around observation and scoring directions (see Box 9).

> **Box 9:** Standardising observations (Pianta and Hamre, 2016, p. 25)
>
> **Training protocol**
>
> - Are there specific directions for learning to use the instrument?
> - Is there a comprehensive training manual or user's guide?
> - Are there videos or transcripts with gold standard scores available that allow for scoring practice?
> - Are other procedures in place that allow for reliability checks (live, via video or via transcript) to ensure that scoring is consistent?
> - Are there guidelines around training to be completed before using the tool?
>
> **Parameters around observation**
>
> - Have these issues been defined?
> - Length of observations
> - Start and stop times of observations
> - Time of day
> - Specific activities to observe
> - Whether observations are announced or unannounced
>
> **Scoring**
>
> - Do users score during the observation itself or after?
> - Is there a predefined observe/score [time] interval?
> - How are scores assigned?
> - Is there a rubric that guides users in matching what they observe with specific scores?
> - Are there examples of the kinds of practices that correspond to different scores?

Archer et al. (2016) is a practical guide for training observers that covers such issues in detail.

### 4.1.3 Further issues

One point mentioned in Box 9 is whether classroom observations for teacher evaluation should be announced or not. Announced observations can result in 'show' lessons where what teachers do may be untypical of their work. For this reason, unannounced visits are believed to provide more realistic evidence of what teachers do. In a

major study of teaching in Latin America and the Caribbean conducted by Bruns and Luque (2015), unannounced visits were carried out in over 15,000 classrooms and the authors claim that 'the evidence is now compelling that the most important element in a robust and meaningful teacher evaluation system is direct observation of classroom practice by trained external observers, preferably on unannounced visits and multiple occasions' (p. 216). In an early paper about teacher evaluation in ELT, Pennington and Young (1989) disagree with this position on the basis that 'unannounced classroom visits are … not only disruptive of the classroom process, but also represent a kind of invasion of privacy' (p. 635). Marzano and Toth (2013) suggest that announced observations, in which teachers are asked to demonstrate specific instructional strategies, allow observers (over several observations) to see a wider range of teaching behaviours (they also believe that when specific teaching strategies are being assessed it is very difficult for teachers to feign competence).

Also mentioned in Box 9 is the use of video for observer training and it is important to note here the role that technology can play in the process of classroom observation. When video recordings of lessons are available digitally, they can be evaluated remotely, which can be more cost-effective than sending observers to schools (though other costs will arise such as for the technology itself and for observer training). Videos can also be reviewed as often as required by observers and provide a record which can be used to examine progress in teaching ability over time. In teacher evaluation in Chile, teachers are required to include a video of one of their lessons in their teaching portfolio (see Section 4.3), while video-based platforms such as Iris Connect are being increasingly used in the UK to support teacher professional development and formative teacher evaluation, including self-evaluation (for a recent review of a project using Iris Connect, see Davies, Perry and Kirkman, 2017).[10] Decisions about the deployment of technology in teacher observation, of course, need to be made with a full understanding of its feasibility in specific contexts.

Finally, it has been argued (see the collection of papers in O'Leary, 2016b), particularly if classroom observation is meant to fulfil a formative function, that a focus on rating discrete skills and behaviours using structured observation tools is counter-productive and that alternative forms of classroom observation are needed. One example is what is called a 'walkthrough'.[11] This is a short, focused observation during which the observer does not need to make notes and is not required to use a checklist (though the observation may focus on particular themes of interest). After the walkthrough the observer can make some notes and use these to support a follow-up reflection discussion with the teacher. As described here (see Stevens, 2016 for an example), walkthroughs allow for the regular collection of information about teaching that can feed into teacher evaluation and which also provides a stimulus for teacher reflection and professional development. This contrast

between highly structured observation rating scales and more open-ended walkthroughs reminds us of the different purposes of teacher evaluation that were discussed earlier: while for summative purposes ratings and grades may be necessary, in a formative context a less structured and formal approach such as a walkthrough may be more productive.

Box 10 summarises key points in relation to the use of classroom observation for teacher evaluation.

---

**Box 10:** Classroom observation

- A key strength of classroom observation is that it provides direct evidence of what teachers do.

- Multiple observations conducted over time by different observers provide a more accurate picture of teaching quality than single observations.

- The reliability of classroom observations can be enhanced through standardisation and rigorous observer training.

- Observation tools should provide a more valid measure of teaching quality if they are aligned with teaching standards.

- Less structured forms of classroom observations such as walkthroughs can contribute to formative teacher evaluation.

- Video-based classroom observations can provide a record of progress over time which can be used for both self-evaluation and remotely by external observers.

---

## 4.2 Student outcomes

Some countries use student outcomes as an indicator of teacher effectiveness; for example, OECD (2015) reported that eight of 19 countries surveyed used student outcomes as part of teacher appraisal, while in an earlier report (OECD, 2013b, p. 34) more specific information about some of these countries was also provided:

> Of the countries surveyed by the OECD, the Slovak Republic reported that student outcomes are used for teacher appraisal at the completion of probation, and Mexico reports that student outcomes (results of standardised assessments) are used for regular appraisal in the context of performance management. Mexico also uses student outcomes to evaluate teacher performance as part of its rewards scheme. In Chile, students' standardised assessment results are used to evaluate groups of teachers (teachers in individual schools) as part of the National Performance Evaluation System (SNED). England, Scotland and Singapore reported that they use student results at some point in the overall teacher evaluation process.

Examples of student learning being linked to teacher evaluation have also been reported from China (Zhang and

---

[10] https://www.irisconnect.com/uk/
[11] See www.pll.asu.edu/p/sites/default/files/lrm/attachments/Reading-The_3-MinuteClassroomWalk-Through%202008.pdf for an overview of Downey's work on walkthroughs.

Ng, 2017) and Japan (Katsuno, 2016), while in the United States, stimulated by the funding available through the 'Race to the Top' initiative, teacher evaluation systems in many different states include a focus on student outcomes (Tucker and Stronge, 2005 discuss various examples).

Test scores are the most obvious form of student outcome and these will be the focus of the discussion below, but it must be noted that other measures of how teaching impacts on students are available. Timperley (2011), for example, talks about student engagement (see also Cinches, Russell, Chavez and Ortiz, 2017; van de Grift et al., 2017) and student well-being, while in the Measures of Effective Teaching project (Ferguson and Danielson, 2014) happiness in class, effort in class, and an increased inspiration to attend college were used (in addition to achievement outcomes) as measures of teaching quality. Changes in students' attitudes towards a subject or in their motivation to study it are further examples of outcomes that can be examined. Clearly, though, the student outcome that has attracted most discussion is achievement, typically as defined by test scores and in particular by test score gains (that is, differences in how students perform before and after instruction).

However, although it may seem obvious that the best way to evaluate teachers is with reference to student achievement, substantial complexities arise in trying to establish direct causal links between what teachers do and what students learn. This is because, **while 'teachers are among the most powerful influences in learning' (Hattie, 2012, p. 22), many factors other than the teacher and outside the teacher's control (such as student characteristics, family background, school attendance, and prior education) affect (in some cases more than teacher effectiveness) what a student learns**. Other variables, such as how students are actually assessed, also have a significant bearing on student outcomes.

One response to such challenges has been the development of what are called 'value-added' models (VAMs). These have been widely discussed in the literature and their value hotly contested (see Braun, 2005 for an accessible introduction). VAMs are complex statistical techniques which analyse student test scores in order to estimate the effect on learning that individual teachers have. Ferguson and Danielson (2014, p. 101) explain that:

> What distinguishes value-added measures from simpler test score growth measures is that they are adjusted for between-classroom differences in student characteristics. Many analysts prefer value added for measuring teacher effectiveness because, if implemented properly, value added approximates a condition in which there is no difference across classrooms in the characteristics of the students. Hence, value added for any particular teacher is an estimate of how much that teacher adds to students' skills and knowledge.

A number of conditions must be met in order for VAMs to be used for teacher evaluation (Goe and Croft, 2009). Firstly, it must be possible to link student test scores to individual teachers. Secondly, individual students' test scores on different subjects over multiple years must be available. Thirdly, the scores must come from standardised tests. One consequence of these requirements is that VAMs can only be used in contexts and for subjects where regular standardised (at least annual) tests take place. This immediately rules out their use in many primary and secondary state education ELT contexts around the world.

As noted above, the use of VAMs in the context of teacher evaluation has been widely debated. On a positive note, Braun (2005, p. 15) makes the following points:

> [VAMs move] the discussion about teacher quality to where it belongs: centered on increasing student learning as the primary goal of teaching. It can also enhance the teacher evaluation process by introducing a quantitative component, as well as by forcing us to re-examine questions of fairness and proper test use. These are major steps in the right direction.

However, concerns about the use of VAMs in teacher evaluation are widespread in the literature and Box 11 summarises some of these (Braun, 2005; Corcoran, 2016; Darling-Hammond, 2013; Grissom, Loeb and Doss, 2016; OECD, 2013a; Raudenbush and Jean, 2014; Steele, Hamilton and Stecher, 2010).

---

**Box 11:** Concerns about using VAMs in teacher evaluation training protocol

- Test scores are affected by a range of variables not directly associated with the teacher.

- VAMs can only be used for subjects that are assessed via standardised achievement tests.

- VAMs do not provide information that teachers can use to improve their teaching.

- Value-added ratings for the same teacher may vary significantly from year to year (especially for teachers receiving very high or very low ratings).

- Differences in teachers' value-added ratings are affected by the composition of their classes.

- VAMs are not reliable measures of broader facets of teachers' work such as contributions to school leadership and building relationships with colleagues.

- VAMs may encourage teachers to teach to the test.

- VAMs require considerable data and technical expertise and are therefore costly to implement.

---

In the context of teacher evaluation, two general conclusions about the use of VAMs, and about student test scores in general, are therefore that (a) they should not be the only measure of teacher quality that is used and (b)

they should not be used for high-stakes decisions about teachers (such as career progression).

This does not mean, of course, that evidence about student learning cannot contribute to teacher evaluation. The key point, though, as Darling-Hammond (2013, p. 88) notes in her discussion of criteria for using measures of student learning in teacher evaluation (see the appendix for these criteria), is that the 'assessment of teachers' contributions to student learning should rely on multiple measures of student learning, not a single test or value-added score'. **Evidence of student learning should thus be obtained in different ways; this can include different kinds of standardised or teacher-designed classroom assessments as well as work presented in other forms, such as student portfolios, that show knowledge growth over time** (see Tucker and Stronge, 2005 for examples).

## 4.3 Teaching portfolios

Teaching portfolios (also referred to as teacher portfolios) are another strategy that can be used in the evaluation of teachers. A teaching portfolio is:

> a collection of materials compiled by teachers to exhibit evidence of their teaching practices, school activities, and student progress … portfolio materials are collected and created by the teacher for the purpose of evaluation and are meant to exhibit exemplary work … The materials gathered are intended to demonstrate fulfillment of certain predetermined standards, and often portfolios are designed to promote teacher reflection and improvement in addition to being used for evaluation. Examples of portfolio materials include teacher lesson plans, schedules, assignments, assessments, student work samples, videos of classroom instruction and interaction, reflective writings, notes from parents, and special awards or recognitions. (Goe, Bell and Little, 2008, p. 30)

Portfolios, therefore, are a very versatile tool. Their reflective component merits particular emphasis here; they are not simply a collection of materials but should demonstrate how teachers have reflected on these and learned from the process. Teaching portfolios have traditionally been print-based, though advances in technology mean that e-portfolios are now also an option (Xerri and Campbell, 2016).

Teaching portfolios can have various benefits. They 'can empower teachers to take charge and have a more active voice in their evaluation' (Attinello, Lare and Waters, 2006, p. 134). **They can provide a holistic, authentic and evidence-based picture of teachers' work over time and thus provide a sound basis for the assessment of teacher competence**. They can stimulate reflection and encourage teachers to take action to address aspects of their work which require improvement. When portfolios provide the basis of professional dialogue among teachers, they can also promote collegiality and collaborative professional learning. Also, portfolios 'can be used with teachers in any subject or grade level and thus are useful

in multiple contexts' (Goe et al., 2008, p. 34).

According to OECD (2013a, 2013b), portfolios are used quite frequently for teacher evaluation, for example, in Singapore, China and Scotland. Teaching portfolios are also used in Chile, where they count for 60 per cent of the total national teacher evaluation score, and Box 12 provides a summary. Taut and Sun (2014) provide further details of how portfolios are used in this context. For example, a call centre is available to answer any logistical questions that teachers have while they are preparing their portfolios. The portfolios are also rated by trained evaluators across eight dimensions of teaching (two of which focus on the quality of teachers' reflections on the lesson). Evaluating the portfolios is a large-scale operation, with around 15,000 portfolios being scored by some 450 raters over a period of four weeks each year.

---

**Box 12:** Teaching portfolios in Chile (Santiago et al., 2013, pp. 49–50)

The portfolio is designed for teachers to provide evidence of their best pedagogical practices. The portfolio is prepared for a given educational level and area of teaching expertise (as defined by the Curriculum and Assessment Unit within the Ministry of Education). Teachers are provided with a Portfolio Manual which, among other things, specifies the descriptors of the Good Teaching Framework which are associated with each of the components of the portfolio. The portfolio consists of two separate modules:

- Set of pedagogical materials (Module 1): the teacher is required to plan and implement an eight-hour teaching unit (providing related materials in writing), to design an end of term assessment for the teaching unit, and to respond to a set of questions about teaching practices (including a reflection on achievements).

- Video recording of a class (Module 2): a 40-minute recording of a regular class together with the completion of a questionnaire about the class. The class is filmed by a cameraman accredited by the Docentemás [national teacher evaluation] team.

---

Another example of a teaching portfolio is the Oregon Teacher Work Sample Methodology (TWSM) (Tucker and Stronge, 2005). The TWSM is a performance-based measure of teacher quality in which teachers compile a portfolio that demonstrates the progress of students, over a period of time, against a set of learning outcomes defined by the teacher. In order to compile the portfolio, teachers work through a series of steps, which include defining the target learning outcomes, planning teaching to address these outcomes, assessing students against these objectives before and after the relevant teaching takes place, and reflecting on and interpreting the results of the process.[12]

---

[12] See www.files.eric.ed.gov/fulltext/ED463282.pdf for a detailed account of the TWSM.

In ELT, the use of portfolios for teacher development (Richards and Farrell, 2005) and for teacher evaluation in higher education contexts (for example, Quirke, 2007; Stoynoff, 2007) has been documented. Published examples of their use for teacher evaluation in primary and secondary state schools are, though, scarce. An exception is Alwan (2007), who discusses the use of portfolios in public school ELT in the United Arab Emirates (UAE). This report describes how, as a supervisor of teachers of English, the author was dissatisfied with the system of annual observations and principal reports that were used to evaluate teachers. In response, she introduced teaching portfolios as a way of obtaining a broader understanding of the work teachers did over a period of time. She also felt that portfolios would give teachers more involvement in the process of their own evaluation. Subsequently, the Ministry of Education in the UAE made portfolios a compulsory component of teacher evaluation. According to Bird and Owais (2004, cited in Alwan, 2007, p. 236) these portfolios served two purposes:

> the first is to improve teaching by encouraging teachers to set forth their own goals, and to self-evaluate these, to encourage professional development practices, to organize records and evidence of and reflection on good practices … The second is to provide an evaluation tool that is formative, fair and balanced.

A number of challenges associated with the use of teaching portfolios in the UAE were identified and these have more general relevance. Teachers were not given sufficient support, particularly in terms of how to write reflections as part of their portfolio. In compiling portfolios, teachers focused on the appearance (for example, size or aesthetic properties) of the portfolio rather than on its quality; the process of compiling the portfolios was not regulated by a structured timetable, meaning that teachers prepared them very near to the submission deadline; teachers received insufficient guidance on what to include in their portfolios; the criteria against which the portfolios would be assessed were not clearly defined; portfolios created tensions and competition among teachers; and there were concerns about the reliability with which the portfolios were evaluated.

Such challenges point to various suggestions for making teaching portfolios an effective tool for teacher evaluation and these are summarised in Box 13 (Alwan, 2007; Gelfer, O' Hara, Krasch and Nguyen, 2015; Goe et al., 2008; OECD, 2013a; Quirke, 2007; Steele et al., 2010).

In some ELT contexts, one additional challenge the use of portfolios may create is linguistic as not all teachers of English will possess sufficient proficiency to write a reflective analysis of their work in English. In such cases it may be feasible for certain parts of the portfolio (for example, where the focus is on teachers' reflections) to be presented in a language other than English.

## 4.4 Student evaluations of teaching

Student evaluations of teaching have been a feature of education contexts around the world for many years and a substantial literature on their use exists (see, for example, Hammonds, Mariano, Ammons and Chambers, 2017; Wachtel, 1998; Zabaleta, 2007). However, as noted by Carlozzi (2017, p. 1), 'despite a near century of study, student evaluations of teaching (SETs) remain contentious', with many arguments both in favour of and against asking students to rate the quality of teaching they receive. Although many of these arguments come from research in university contexts (for an example from ELT, see Burden and Troudi, 2007), they are largely relevant to teacher evaluation in primary and secondary state school contexts and will be summarised below.

Various benefits of formal SETs have been identified, such as that they:

- are an important source of feedback to teachers
- can contribute to improving the quality of teaching
- allow student voices to be heard
- are an efficient way of collecting student feedback.

However, many concerns have also been raised about SETs. Benton and Ryalls (2016), though, suggest that many of these concerns are in fact based on a number of misconceptions, which are listed in Box 14. In response to each of these, the authors present evidence to support their argument that SETs are in fact a valuable source of feedback on teaching effectiveness. Citing Darwin (2012), Hammonds et al. (2017, p. 31) also conclude that 'research conducted on SETs over the last 30 years suggests that SETs, while they do have shortcomings, provide valuable information regarding teaching effectiveness'.

One particular shortcoming in the use of SETs relates to the design of the instruments. Poorly designed tools are in fact

---

**Box 13:** Using teaching portfolios in teacher evaluation

- Provide clear guidelines about the purpose, format, length and content of the portfolio.

- Specify how portfolios will be assessed, including the assessment criteria.

- Provide adequate training for evaluators.

- Define the weight that portfolios carry in the teacher evaluation process.

- Ensure that teaching portfolios have a formative not just summative purpose.

- Give teachers the support they need, particularly with the reflective component of the portfolio.

- Use portfolios to document what teachers typically do rather than act as products created especially for the purposes of teacher evaluation.

- Ensure that the time and costs associated with the completion and evaluation of the portfolios are commensurate with the resources available.

recognised as a major cause of unreliability in SETS and one consistent message from the literature is that if student evaluations are to make a useful contribution to teacher evaluation it is important that high-quality instruments are used. In particular, Benton and Ryalls (2016) argue that students should only be asked to evaluate aspects of teaching which they are qualified to judge. In this respect, Goe, Bell and Little (2008, p. 41) note that 'students are not usually qualified to rate teachers on curriculum, classroom management, content knowledge, collegiality, or other areas associated with effective teaching'. **Aspects of teaching which students are considered qualified to judge are, among others, teacher enthusiasm, course organisation, fairness of examinations, teacher rapport with students and clarity of teachers' explanations** (for further discussion, see Benton and Cashin, 2012).

---

**Box 14:** Misconceptions about student evaluations of teaching (Benton and Ryalls, 2016)

- Bad teachers get better evaluations.

- Demanding teachers receive lower evaluations.

- Students are not qualified to judge teaching effectiveness.

- Student evaluations are not reliable.

- Personal factors unrelated to learning influence ratings.

- Students tend to be motivated more by anger about a low grade than satisfaction.[13]

- Millennial students are more punishing in their ratings.

---

Whereas the use of SETs in higher education is widespread, OECD (2013b) notes that, in primary and secondary contexts, student feedback does not play a major role in the evaluation of teachers. This does not mean that in OECD countries teachers do not collect feedback from students, but that the results are often used only by the teachers to improve their teaching and not examined formally in any evaluation of teacher effectiveness. In China, though, student feedback is widely used in teacher evaluation and it has been a topic of extensive discussion. Liu and Teddlie (2005) summarise some research on the issue and cite a study by Wang and Sui (2005) which found that, across 10 schools, principals and teachers were in favour of using student evaluations. This study, though, also highlighted various concerns including teachers trying to please students in order to get positive ratings. Such issues can be exacerbated in contexts, such as China, where teacher evaluation is linked to pay (Liu and Zhao, 2013). In fact, it is suggested that, for summative teacher evaluation, student evaluations should not carry significant weight (Benton and Ryalls, 2016 suggest 30–50 per cent) and should, as already noted, be but one of several ways of assessing teaching effectiveness.

The value of SETs is also discussed by Ferguson (2012), whose Tripod tool was extensively evaluated in primary and secondary schools in the USA as part of the Measures of Effective Teaching project (for a discussion, see Ferguson and Danielson, 2014). Tripod focuses on three domains of effective teaching – content, pedagogy and relationships – and examines students' perceptions of these in relation to seven areas referred to as the '7 Cs': challenge, control, care, confer, captivate, clarify and consolidate (for further details, see Ferguson, 2012). Analyses of Tripod indicate that it does produce valid and reliable measures of teaching quality (Polikoff, 2015; Wallace, Kelcey and Ruzek, 2016). In line with the conclusions reached above by Benton and Ryalls (2016), work with Tripod suggests that SETs – when well-designed – can support teacher evaluation. As Ferguson (2012, p. 27) concludes:

*Doubts about whether student responses can be reliable, valid, and stable over time at the classroom level are being put to rest. We are learning that well constructed classroom-level student surveys are a low burden and high-potential mechanism for incorporating students' voices in massive numbers into our efforts to improve teaching and learning.*

SETs can be administered online (including via mobile devices) as well as in paper format. Treischl and Wolbring (2017) provide an analysis of how delivery mode may affect student responses; this indicates that response rates to online SETs tend to be lower but that, in terms of reliability, the research available suggests that there are no significant differences between paper and online SETs. The points made above regarding the importance of careful instrument design, of course, apply equally to online SETs.

One final point to make here is that the decision to introduce SETs for the purposes of teacher evaluation needs to take into account teachers' and students' prior experiences of evaluating teaching. For example, if students have never been asked to provide feedback on their teachers, they may initially be unsure as to why they are being asked to do so. This is even more likely in contexts where teachers are highly respected and their competence is not normally subjected to scrutiny. The process also has the potential to affect the relationship between students and teachers, especially when teachers are concerned about the consequences of student feedback. Younger children are also often keen to make their teacher happy and this will influence how they respond to certain questions about their teachers. Such concerns highlight the potentially sensitive nature of SETs and the need for their introduction to be grounded in an appropriate analysis of the target socio-educational context.

---

[13] The claim here is that dissatisfied students are more likely to complete student evaluations of teaching than those that are satisfied.

Box 15 summarises key points regarding the use of student evaluations of teaching in teacher evaluation.

<div style="border: box">

**Box 15:** Student evaluations of teaching (SETs)

■ Most research on SETs has been done in university contexts.

■ In some education systems SETs are used in teacher evaluation, but in many they are an informal source of feedback to the teacher.

■ SETs can produce valid assessments of teaching quality, but only when well-designed instruments are used.

■ Many common objections to SETs can be addressed when they are appropriately designed, administered, interpreted and used.

■ In teacher evaluation, SETs should be used alongside other measures.

■ SETs should also be administered on multiple occasions over time.

■ SETs can make both a formative and summative contribution to teacher evaluation.

■ In high-stakes teacher evaluation contexts, SETs should not carry too much weight.

■ SETs need to take into account the socio-educational context in which they are to be used.

</div>

## 4.5 Self-evaluation

Self-evaluation is the process through which individuals assess their own competence. In teaching, it is recognised as an important part of a teacher's professional development and can also contribute to teacher evaluation. Many examples are available of tools and frameworks that support teachers in the process of self-evaluation. The General Teaching Council for Scotland, for example, offers teachers a 'self-evaluation wheel' where teachers can assess themselves against eight aspects of their work.[14] Self-evaluation is more formally a part of teacher evaluation in various countries. In Kosovo, for example, it is proposed that self-assessment contributes 10 per cent of the overall teacher evaluation score.[15] In Chile, self-assessment is used as follows:

*The self-evaluation consists of a structured questionnaire organised according to the four domains of the Good Teaching Framework (GTF). Its objective is to generate teachers' reflection of their own practice and encourage teachers to review the GTF. The self-evaluation proposes 12 areas (3 areas per GTF domain), each related to a specific criterion*

[14] www.gtcs.org.uk/professional-standards/self-evaluation/self-evaluation-standard-full-registration.aspx
[15] www.ec.europa.eu/dgs/education_culture/repository/education/events/2016/wbp/teacher-evaluation-tools-criteria_en.pdf

*in the GTF, on which the teacher rates his or her performance in four possible levels: Unsatisfactory, Basic, Competent, Outstanding. There are no open-ended questions. Teachers also have the possibility of adding information about the context for their teaching. Teachers are given guidelines with a protocol to rate themselves. (Santiago et al., 2013, p. 46)*

Portugal also uses self-assessment as part of teacher evaluation. In this case:

*teachers are required to submit annually a self-appraisal report in which they reflect on their practices in the following areas: teaching; activities promoted; analysis of results obtained; contribution to the objectives and goals set in the educational project of the school; and professional development undertaken and its contribution to their work. The self-evaluation report cannot exceed three pages. (Santiago, Donaldson, Looney and Nusche, 2012, p. 71)*

Teacher self-evaluation, then, as these examples show, can take different forms, from more structured questionnaires to more open-ended reports. Teaching portfolios (see Section 4.3) also contain an element of self-evaluation because teachers are required not only to select and assemble evidence of their work but also to reflect on its effectiveness.

In ELT, as discussed in Section 3.3, various frameworks are also available which teachers can use to self-assess their current level of competence or development. While organisations such as Cambridge Assessment English and the British Council (which has developed a structured self-assessment tool for teachers – see Borg and Edmett, forthcoming) make wide use of the frameworks they have developed in their ELT teacher development work, published research into the design of such frameworks and of how teachers use them for self-assessment remains limited.

Overall, though, there is general consensus in the literature that **self-evaluation is a desirable feature of teacher evaluation. It gives teachers a greater sense of involvement in the evaluation process and allocating responsibility to teachers in this way acknowledges their status as professionals** (Pennington and Young, 1989). This can have a positive impact on teacher motivation (for a recent analysis from Africa, see UNESCO-IICBA, 2017). Also, because it is informed by teachers' knowledge of what they typically do, self-evaluation can provide a more valid sense of teacher competence than one-off classroom observations can (Marzano and Toth, 2013).

One concern associated with self-evaluation of any kind, though, relates to the extent to which individuals are able to provide an accurate assessment of their own competence. Research with health professionals suggests that they are limited in their ability to self-assess accurately, particularly those who are less skilled and those who are most confident (Davis et al., 2006). Relevant evidence from

teaching is limited, though an analysis of the self-evaluation component of teacher evaluation in Chile concluded that score inflation did take place and that this was unsurprising given the high-stakes nature of the evaluation. It was recommended, therefore, that 'self-assessment should serve exclusively formative purposes' (Taut and Sun, 2014, p. 23). **Teacher self-assessment is thus more likely to generate accurate results where the focus is on using these to inform professional development rather than for accountability**. Concerns about the validity of self-assessments may also explain the relatively modest weight it carries in teacher evaluation systems where it is utilised (see Box 7).

Teachers' previous experience of self-evaluation should be considered before it is introduced as a teacher evaluation strategy. If teachers have not had previous opportunities to engage in self-evaluation, are unfamiliar with the concept and its assumptions, and work in an educational system where self-assessment is not something that students do either, **considerable preparatory work will be required to create the conditions where teacher self-evaluation can function effectively**. This is a further reminder that decisions about the way teachers are evaluated will always need to be situated within broader considerations of teaching and learning in specific contexts.

## 4.6 Teacher tests

As Marzano and Toth (2013) note, a complex skill such as teaching requires a sound cognitive (i.e. knowledge) base; using teacher tests to assess what teachers know is thus a relevant component of teacher evaluation. Testing prospective teachers as a means of gaining a licence to teach is fairly common; in many contexts this occurs at university, while in others, there are additional national licensing examinations that must be passed. In the USA, teacher certification requirements vary across states but typically involve some form of testing (see, for example, the Praxis tests of teacher knowledge developed by ETS).[16] The use of tests with practising teachers, though, is much less common and, as I discuss below, can be controversial.

In any kind of teacher test, decisions are required about the focus of the test; for example, will it assess knowledge of the subject or knowledge of teaching and learning? For the former, in ELT, various tests of English proficiency for teachers are available (for example, the British Council's Aptis for Teachers[17]) while in terms of pedagogical knowledge an examination such as the Teaching Knowledge Test (TKT) offered by Cambridge Assessment English examines what teachers know about language, language learning and teaching, lesson planning and use of resources, and managing the teaching and learning process.[18] A second decision in teacher tests relates to test format; for example, the TKT is a multiple-choice test, but it must be noted that:

---

[16] www.ets.org/praxis/about

[17] www.britishcouncil.org/school-resources/aptis/teachers

[18] www.cambridgeenglish.org/teaching-english/teaching-qualifications/tkt/

[19] See www.gppreview.com/2016/03/10/mexicos-education-reform-what-went-wrong/ for a perspective on the educational reforms in Mexico.

*although some aspects of teacher knowledge, such as factual knowledge of content, may be adequately assessed with multiple-choice questions, the assessment of critical understanding is likely to require tests on which examinees produce extended responses rather than selecting among prespecified phrases. The assessment of teacher judgment and decision making will require testing formats in which a richer context can be presented for the decision to be taken. (Haertel, 1991, pp. 12–13)*

Teacher performance assessments provide an alternative to multiple-choice tests. In these, for example, teachers are required to produce a lesson plan based on specific parameters, or to analyse and respond to examples of students' work. Although such tests may be seen as a more authentic measure of what teachers can do, and also provide access to their thinking, they are much costlier, in terms of administration and assessment (including evaluator training), than those which use multiple-choice formats.

Apart from the conceptual and technical issues discussed so far, asking practising teachers to take tests and using these in teacher evaluation can be a very sensitive matter. In many contexts, it is assumed that graduating from a university is evidence that teachers have a sufficient level of subject matter and pedagogical knowledge for teaching. The use of tests with practising teachers challenges this assumption and can raise objections both from universities (who may feel that the quality of the initial teaching qualifications they provide is being questioned) and from teachers themselves. It is, therefore, essential that any decision to use teacher tests in the process of teacher evaluation be taken very carefully. Teachers may also be concerned about how test results will be used and who will have access to them. In contexts where teacher unions are powerful, proposals to introduce teacher testing can also become highly politicised, as the example of Mexico below illustrates.

OECD (2013b) does not identify many countries where teacher tests are used for teacher evaluation. Mexico is one exception though. In 2013, as part of a broader education reform, a new system of compulsory teacher performance tests was announced. This initiative has been characterised by substantial opposition from Mexico's powerful teacher unions, including teacher protests. One major source of opposition to the teacher tests is that they are seen to be punitive, given that teachers who cannot pass (they get more than one chance) could lose their job, and they are seen as an attempt to impose more state control on teachers. There have also been complaints about the validity of the tests (in terms of whether they actually assess teaching skills).

The Mexican experience is a clear example of the conflicts that can arise when testing is introduced for teacher evaluation in a context where the conditions are not immediately amenable to such an approach to assessing teacher competence or, perhaps, to summative teacher evaluation more generally.[19] Mexico also illustrates very clearly the point made by Bruns and Luque (2015, p. 3) in

their analysis of teacher evaluation in Latin America: 'The deepest challenge in raising teacher quality is not fiscal or technical, but political, because teachers' unions in every country in Latin America are large and politically active stakeholders'. A recent global analysis of education (World Bank, 2018) cites further examples from South Africa and Peru where teacher evaluation reform was opposed. In Peru, for example, in response to a legal challenge by unions, a new law requiring all teachers to undergo evaluations was revised so that it only applied to newly hired teachers. This does not imply, of course, that unions will always impede efforts to enhance teacher quality; they can be powerful allies and in some cases resistance to reform may also be justified because it ignores many of the principles for effective teacher evaluation highlighted in this review.

## 4.7 Other strategies

Three further final strategies for teacher evaluation are discussed briefly in this section. While these are noted in the literature, there is less global evidence of them than the more commonly-used strategies discussed above.

### 4.7.1 Professional conversations

Teacher evaluation may also involve teachers in the process of professional conversations about their work. These may take the form of structured interviews where teachers talk to the school leader about some aspect of their teaching; for example, as OECD (2013b) notes, teachers may be required to talk about previously established targets for student learning and the process they have been through in addressing these. Such a process may be linked to some other component of teacher evaluation, such as a teaching portfolio (see Section 4.3). In some countries where teacher evaluation systems are not regulated by national frameworks (such as Sweden, Finland and Denmark[20]), individual professional dialogues between teachers and the school leaders are the main form of feedback for teachers (Stewart, 2013). While conversations or interviews with teachers allow for an in-depth discussion of their work, they are time-consuming to conduct (Campbell et al., 2004). Also, as Isoré (2009, p. 13) notes, 'teachers' propensity to reveal their real weaknesses and fears during interviews depends on their confidence in the interviewer and their perceptions of the possibility to receive relevant and constructive feedback from the evaluation process'.

### 4.7.2 Peer evaluation

Peer evaluation is the process through which a teacher is assessed by a colleague rather than by a school leader, line manager or external evaluator. Although peer evaluation does not seem too common in teacher evaluation, there are examples in the literature of how it can be used. For example, as noted in Section 4.1, the classroom observation of teachers can be conducted by peers (in which case it becomes peer observation –

see, for example, Cosh, 1999). Speer and Harich (2005) refer to the introduction of peer observation in Germany, including a project called Eiver where teachers from different schools visited one another for the purposes of peer evaluation. According to the Eiver website,[21] the project showed that 'the critical feedback of the peers is an important complement to the internal evaluation of school development projects and contributes effectively to the quality assurance in the schools'. In the Netherlands, peer evaluation takes place through an initiative called SKOOP:

> which involves teachers and staff from one school visiting and reviewing their counterparts in another. The methodology is based on a supervisory framework developed by the Dutch inspectorate of education. A department or team of tuition or management staff visits another school where they observe lessons and conduct panel interviews in order to form an impression of the school. That impression is then discussed with the school authorities and a written report is produced. The results to date are encouraging, and it is believed that this type of peer review can make a valuable contribution to the professionalisation of teaching staff, helping them to make full use of their professional autonomy. (OECD, 2013b, p. 63)

Another example of peer evaluation is known as peer assistance and review (PAR). Darling-Hammond (2013) describes how PAR is used in the USA and concludes that it has been effective in evaluating and supporting teachers (either novices or veterans who are struggling). PAR involves the use of trained mentors who, together with administrators, contribute to the collection and review of evidence about teacher competence and make recommendations about the outcome of the teacher evaluation process. In the case of struggling teachers, this may include recommending that a teacher's contract not be renewed.

**Peer evaluation is based on the belief that teachers have the knowledge to assess the competence of their peers; it is important, therefore, that teachers receive training for peer evaluation and that, additionally, it is suitably resourced**. For example, in PAR programmes in the USA, the mentors have specific workload time allocated for peer review and in some cases mentoring is a full-time position. If teachers do not feel that the demands of peer evaluation are realistic, their engagement in the process may be low, as, for example, in a recent peer observation project in the UK (Worth, Sizmur, Walker, Bradshaw and Styles, 2017) where teachers who were expected to complete 12 observations a year only completed six in two years.

### 4.7.3 Parent feedback

'Because parents are part of the work of teachers, and see it from an important and unique point of view, there is a need to somehow incorporate parent perceptions into teacher evaluation systems in order to get the broadest look at teacher quality' (Peterson, Wahlquist, Brown and

---

[20] Finland's case illustrates how an effective education system (as measured on international comparisons of student performance such as PISA) is possible without a formal national teacher evaluation framework (but this does not mean, of course, that teacher evaluation does not occur).

[21] www.teamlearn.de/LotusQuickr/b-1-eiver/Main.nsf/h_Toc/4df38292d748069d05256708001 67212/?OpenDocument

Mukhopadhyay, 2003, p. 308). How precisely this can be achieved is a matter of debate, but, as in the discussion of student evaluations above (Section 4.4), it is important to ensure that parents are asked appropriate questions and that instruments are carefully designed.[22] Also, parent feedback should not carry excessive weight in the overall process of teacher evaluation.

In Korea, 'parents provide information on satisfaction levels by filling in questionnaires composed of checklist questions and descriptive opinion writing' (Kim et al., 2010, p. 54). The same report notes that to improve the quality of parent evaluations of teachers, training is offered to parents by individual schools. Another example of parent involvement in teacher evaluation comes from Italy, where, as part of an experiment with 33 schools, parents completed a questionnaire in which they were asked to nominate teachers who they felt were highly respected for their professional behaviour. Parent input was used together with that from students and peers to reach decisions about which teachers were most highly respected in each school (for further details, see OECD, 2013b).

Overall, though, how parent feedback can be effectively integrated into teacher evaluation has not been widely studied. In addition to technical issues related to the design and administration of tools for parents and other factors which affect how parents respond (the age of their children seems to be a factor), involving parents in teacher evaluation also has resource implications (i.e. the process can be costly) and raises various interpersonal and political concerns too (for example, teachers may feel threatened when parents are asked to judge their quality). Peterson et al. (2003) discuss such issues. They also make the important point that **positive parent evaluations of teaching cannot be equated with teacher effectiveness**. **Rather, they provide insight into parental perceptions of teachers which can contribute to the broader process of teacher evaluation.**

---

[22] See www.connecticutseed.org/wp-content/uploads/2012/12/Parent_Survey_Question_Bank.pdf, though this list extends beyond a specific focus on teacher quality.

# 5

# Recommendations

As noted by Bruns and Luque (2015, p. 35), among others, 'top education systems invest heavily in the evaluation of teacher performance' and it is widely acknowledged in the literature that teacher evaluation is an important component of broader national or state-level policies for teachers. This review has drawn on this literature to identify the elements of effective teacher evaluation systems more generally and discuss in detail a range of specific ways in which evidence of teacher quality can be obtained. To conclude, the following recommendations for the evaluation of teachers in state school ELT contexts are made. Throughout these recommendations, 'institution' refers to educational organisations (which may consist of one or more establishments such as schools and colleges) and to state- or national-level education systems more generally.

1   The quality of teachers has a significant impact on what students learn and teacher evaluation is a key element in the process of improving teaching. Institutions wanting to enhance the quality of student learning should, therefore, examine how teacher evaluation is currently approached and how it might be improved.

2   A framework for teacher evaluation should define, in a transparent way, basic parameters such as who is to be evaluated, for what purpose, against which criteria, how often, who the evaluators will be, how evidence will be collected and analysed, and how the results will be used.

3   Overall, the literature discussed in this review indicates that effective teacher evaluation is a multi-dimensional process which acknowledges the complexity of teaching, employs a range of good-quality measures, utilises input from different stakeholders (including teachers themselves), facilitates fair decisions, gives teachers appropriate levels of support, and contributes to teacher professional development and to improving student outcomes.

4   All teachers (and not just novices or those experiencing problems) can benefit from teacher evaluation, but effective teacher evaluation systems will offer differentiated opportunities for teachers at different stages of their careers.

5   Both the formative and summative purposes of teacher education are important. It is vital, though, that institutions recognise the differences between them and approach teacher evaluation with these differences in mind. Overall, though, institutions should ensure that teacher evaluation gives teachers opportunities to improve rather than only delivering summative measures of teacher quality (such as scores or grades).

6   At the same time, it is important to stress that teacher evaluation can play an important role in identifying teachers who are not performing sufficiently well. An effective teacher evaluation policy will specify the steps an institution will take to support such teachers so that they have an opportunity to achieve the required standards.

7   The involvement of teachers in the design and implementation of teacher evaluation systems is highly desirable. Institutions should create meaningful opportunities for teacher input and feedback to ensure they feel sufficiently involved in the teacher evaluation process.

8   Institutions also need to be aware that teacher evaluation will not always be received positively by teachers; some will feel threatened by the process or consider it to be unfair. In consulting teachers, institutions should be sensitive to such concerns, seek to understand them, and respond in a considered manner.

9   Teacher evaluation should also engage all other relevant stakeholders (in addition to teachers) in a consultative process. Especially for larger-scale reform, the support of groups such as school leaders, teacher associations and teacher unions will impact positively on the extent to which teacher evaluation reform succeeds.

10  Teacher evaluation should be underpinned by a clear vision of what competent teaching is. Institutions thus need to define ELT teacher competence and to use this as a reference point when the quality of teachers is being assessed. Various ELT-specific frameworks that can assist with this task are now available. These cover a range of competences related to teachers' English proficiency, their declarative knowledge about language, theoretical knowledge about teaching, practical planning and teaching skills, and reflective capacities.

11  While teacher competency frameworks that are not subject-specific provide a useful resource, and having the same criteria for teachers of different subjects may be more administratively efficient, ELT teacher evaluation is likely to be more meaningful for teachers and productive for professional development when it draws on a core set of subject-specific criteria.

12  Teachers' English proficiency is a fundamental component of teacher quality in ELT and the assessment of teachers' proficiency is, therefore, a warranted focus for teacher evaluation. However, it is recommended that institutions approach this issue sensitively, given concerns that teachers and other stakeholders may have both about teacher testing

generally and specifically about testing teachers' English. It is also important that institutions make principled decisions about the kind and level of English proficiency teachers require and how best to assess it.

13   Teacher quality extends beyond classroom performance. Teacher evaluation systems should acknowledge this broader view of teaching by examining not just what teachers do in the classroom and how it affects students, but also teachers' professional activities more generally, their contributions to the school, and how they collaborate with other teachers.

14   Many different ways of assessing teacher quality exist and teacher evaluation should draw on multiple sources of evidence over time and from different evaluators. No one single source of evidence can capture teacher quality in a sufficiently rich way.

15   One-off classroom observations have traditionally been given disproportionate weight in the process of evaluating teachers. It is important to recognise the severe limitations of such an approach and to work towards ways of evaluating teachers that provide a more accurate and broader picture of their competence.

16   Classroom observations of teachers have also traditionally been characterised by low levels of validity and reliability. Much contemporary knowledge of how to improve the quality of classroom observations (through high-quality standardised tools and observer training) is, however, now available and it is recommended that institutions draw on this knowledge to improve the way they use observations in the process of teacher evaluation.

17   It is recommended that teacher evaluation involve a range of evaluators. These may be internal to an institution or external to it. Internal evaluators may include teachers themselves (self-evaluation), school leaders, peers and students; external evaluators may include inspectors or supervisors, teachers from other schools and parents.

18   It is also essential that institutions provide the training required for evaluators to fulfil their responsibilities effectively. Inconsistent or uninformed evaluations of teachers will have minimal value and also be a source of dissatisfaction among teachers.

19   Teacher evaluation should be driven by a desire to improve student learning, and student outcomes do have an important role to play in the evaluation of teachers. Achievement is only one possible student outcome, though, and it is recommended that institutions consider others such as student satisfaction, happiness and motivation. Also, while value-added models which link teacher effectiveness to student achievement have been widely promoted, serious reservations about their use have also been noted and institutions should give these due consideration.

20   Students can contribute to teacher evaluation but it is important that institutions involve students appropriately and, in particular, make use of well-designed tools for collecting student feedback on the quality of teachers. Institutions should also assess the contextual appropriateness of asking students to evaluate their teachers and provide the support that might be needed for both students (who may have never been asked to evaluate their teachers before) and teachers (who may be concerned about what students will say and how the results will be used). With children, the way evaluations of teaching are elicited and interpreted should take the characteristics of young learners into account.

21   Teacher self-evaluation, especially when it is formative, should also be a component of teacher evaluation because it can give teachers a positive sense of responsibility for the evaluation of their own work and contribute to a more holistic assessment of what teachers can do. In summative contexts, it is recommended that self-evaluation not be given much weight due to the natural tendency in such contexts for teachers to inflate their self-ratings.

22   While the formative nature of teacher self-evaluation means that some imprecision in how teachers rate their own competence can be tolerated, in contexts where self-evaluation is a novel practice, teachers will nonetheless benefit from support (such as training in how to self-assess realistically) that allows them to engage meaningfully in the process.

23   While teachers may be asked to produce work especially for the purposes of teacher evaluation, institutions should seek as far as possible to link teacher evaluation to the work teachers normally do. This reduces the tendency in teachers to showcase work that is not typical of their teaching or that has no connection with what students are learning. Portfolios can be very useful in this regard, especially when they focus on lessons that teachers have actually taught and work that students produced.

24   Portfolios and other forms of teacher evaluation require teachers to reflect on their work. It is vital, though, that institutions give teachers the support they need to understand what reflection entails. As with self-evaluation, teachers should not be asked to reflect on their work unless they have been given opportunities to develop the skills to do so.

25   Conventional approaches to teacher evaluation have been limited by their inability to differentiate between teachers with very different levels of competence. It is important, therefore, that institutions adopt teacher evaluation systems which do discriminate between teachers; this allows teachers who are doing well to receive recognition and those who need to improve to be identified and given support.

26   Teacher evaluation can have significant resource implications and, in making decisions about how to evaluate teachers, institutions will need to balance what is theoretically desirable (such as multiple sources of evidence about teacher quality) with what is practically feasible. The outcome of this process,

though, must still remain faithful to the basic tenets of effective teacher evaluation.

27   A serious approach to teacher evaluation must be underpinned by a concern for quality at all levels of an institution. Teacher evaluation should thus ideally occur alongside equivalent attention to the quality of learning and the quality of educational leadership.

28   Evaluator training is an essential component of a teacher evaluation system. Effective teacher evaluation is not possible unless the evaluators (school leaders, inspectors, teachers, students and even parents) possess the knowledge and skills they need to contribute in an appropriate manner to the process of assessing teacher competence.

29   One final recommendation for the field of ELT more generally is that much more research is required into how teacher evaluation operates in practice. In particular, case studies of ELT teacher evaluation in specific contexts and of how teachers experience it are required to illustrate how theoretical and practical concerns shape actual practices. Such insights would have significant implications for the development of teacher evaluation in ELT.

# References

Abdelaziz, W, Geberi, N E, Etshim, R, Haitova, N, Alvarez, F J H D, Larsen, J, . . . Wilson, L (2016) *Case studies of teacher evaluation systems around the world*. Kalamazoo, MI: Western Michigan University.

Alwan, F (2007) Teacher voice in teacher evaluation: Teaching portfolios in the United Arab Emirates, in Coombe, C, Al-Hamly, M, Davidson, P and Troudi, S (eds) *Evaluating teacher effectiveness in ESL/EFL contexts* (pp. 232–244). Ann Arbor, USA: University of Michigan Press.

Archer, J, Cantrell, S, Holtzman, S L, Joe, J N, Tocci, C M and Wood, J (2016) *Better feedback for better teaching: A practical guide to improving classroom observations*. San Francisco, CA: Jossey-Bass.

Archer, J, Kerr, K A and Pianta, R C (2014) Why measure effective teaching?, in Kane, T, Kerr, K A and Pianta, R C (eds) *Designing teacher evaluation systems: New guidance from the measures of effective teaching project* (pp. 1–5). San Francisco, CA: John Wiley & Sons.

Attinello, J R, Lare, D and Waters, F (2006) The value of teacher portfolios for evaluation and professional growth. *NASSP Bulletin, 90*(2), 132–152.

Avalos, B and Assael, J (2006) Moving from resistance to agreement: The case of the Chilean teacher performance evaluation. *International Journal of Educational Research*, 45(4), 254–266.

Bell, C A, Qi, Y, Croft, A J, Leusner, D, McCaffrey, D F, Gitomer, D H and Pianta, R C (2014) Improving observational score quality: Challenges in observer thinking, in Kane, T, Kerr, K A  and Pianta, R C (eds) *Designing teacher evaluation systems: New guidance from the measures of effective teaching project* (pp. 50–97). San Francisco, CA: John Wiley & Sons.

Benton, S L and Cashin, W E (2012) *Student ratings of teaching: A summary of research and literature*. Idea paper# 50. IDEA Center, Inc.

Benton, S L and Ryalls, K R (2016) *Challenging misconceptions about student ratings of instruction*. Idea paper# 58. IDEA Center, Inc.

Borg, S. and Edmett, A. (forthcoming) Developing a self-assessment tool for English language teachers. *Language Teaching Research.*

Braun, H I (2005) *Using student progress to evaluate teachers: A primer on value-added models*. Princeton, NJ: Educational Testing Service.

Bruns, B and Luque, J (2015) *Great teachers: How to raise student learning in Latin America and the Caribbean*. Washington, DC: World Bank Publications.

Burden, P and Troudi, S (2007) An evaluation of student ratings of teaching in a Japanese university context, in Coombe, C, Al-Hamly, M, Davidson, P and  Troudi, S (eds) *Evaluating teacher effectiveness in ESL/EFL contexts* (pp. 152–166). Ann Arbor, USA: University of Michigan Press.

Campbell, J, Kyriakides, L, Muijs, D and Robinson, W (2004) *Assessing teacher effectiveness: Developing a differentiated model.* London, England: Routledge.

Carlozzi, M (2017) Rate my attitude: Research agendas and ratemyprofessor scores. *Assessment & Evaluation in Higher Education*, 1–10.

Cinches, M F C, Russell, R L V, Chavez, J C and Ortiz, R O (2017) Student engagement: Defining teacher effectiveness and teacher engagement. *Journal of Institutional Research South East Asia, 15*(1), 5–19.

Coombe, C, Al-Hamly, M, Davidson, P and Troudi, S (eds) (2007) *Evaluating teacher effectiveness in ESL/EFL contexts*. Ann Arbor, USA: University of Michigan Press.

Corcoran, S P (2016) Potential pitfalls in the use of teacher value-added data, in Grissom, J A and Youngs, P (eds) *Improving teacher evaluation systems: Making the most of multiple measures* (pp. 51–62). New York, NY: Teachers College Press.

Cosh, J (1999) Peer observation: A reflective model. *ELT Journal*, 53(1), 22–27.

Council of Chief State School Officers (2013) *InTASC model core teaching standards and learning progressions for teachers 1.0*. Available online at: www.ccsso.org/Documents/2013/2013_INTASC_Learning_Progressions_for_Teachers.pdf

Danielson, C and McGreal, T L (2000) *Teacher evaluation to enhance professional practice*. Alexandria, VA: ASCD & ETS.

Darling-Hammond, L (2013) *Getting teacher evaluation right: What really matters for effectiveness and improvement.* New York, NY: Teachers College.

Davies, P, Perry, T and Kirkman, J (2017) *Iris Connect: Developing classroom dialogue and formative feedback through collective video reflection*. London, England: Education Endowment Foundation.

Davis, D A, Mazmanian, P E, Fordis, M, Van Harrison, R, Thorpe, K E and Perrier, L (2006) Accuracy of physician self-assessment compared with observed measures of competence: A systematic review. *Journal of the American Medical Association, 296*(9), 1094–1102.

De Jager, T, Coetzee, M J, Maulana, R, Helms-Lorenz, M and van de Grift, W (2017) Profile of South African secondary-school teachers' teaching quality: Evaluation of teaching practices using an observation instrument. *Educational Studies, 43*(4), 410–429.

Department for Education, UK (2012) *Teacher appraisal and capability: A model policy for schools*. Available online at: www.gov.uk/government/publications/teacher-appraisal-and-capability-model-policy

Dunkin, M J (1997) Assessing teachers' effectiveness. *Issues in Educational Research, 7*(1), 37–51.

Ferguson, R F (2012) Can student surveys measure teaching quality? *Phi Delta Kappan, 94*(3), 24–28.

Ferguson, R F and Danielson, C (2014) How framework for teaching and Tripod 7 Cs evidence distinguish key components of effective teaching, in Kane, T, Kerr, K A and Pianta, R C (eds) *Designing teacher evaluation systems: New guidance from the measures of effective teaching project* (pp. 98–143). San Francisco, CA: John Wiley & Sons.

Freeman, D, Katz, A, Garcia Gomez, P and Burns, A (2015) English-for-teaching: Rethinking teacher proficiency in the classroom. *ELT Journal, 69*(2), 129–139.

Gargani, J and Strong, M (2014) Can we identify a successful teacher better, faster, and cheaper? Evidence for innovating teacher observation systems. *Journal of Teacher Education, 65*(5), 389–401.

Gelfer, J, O' Hara, K, Krasch, D and Nguyen, N (2015) Teacher portfolios: An effective way to assess teacher performance and enhance learning. *Early Child Development and Care, 185*(9), 1495–1503.

Glazerman, S, Goldhaber, D, Loeb, S, Raudenbush, S, Staiger, D O, Whitehurst, G J and Croft, M (2011) *Passing muster: Evaluating teacher evaluation systems*. Washington, DC: Brooking Institution.

Goe, L, Bell, C and Little, O (2008) *Approaches to evaluating teacher effectiveness: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality.

Goe, L and Croft, A (2009) *Methods of evaluating teacher effectiveness*. Washington, DC: National Comprehensive Center for Teacher Quality.

Good, T L and Lavigne, A L (2015) Rating teachers cheaper, faster, and better: Not so fast. *Journal of Teacher Education, 66*(3), 288–293.

Goodwyn, A (2017) *Expert teachers: An international perspective*. London, England: Taylor & Francis.

Grissom, J A, Loeb, S and Doss, C (2016) The multiple dimensions of teacher quality: Does added-value capture teachers' nonachievement contributions to their schools?, in Grissom, J A and Youngs, P (eds) *Improving teacher evaluation systems: Making the most of multiple measures* (pp. 37–50). New York, NY: Teachers College Press.

Grissom, J A and Youngs, P (2016) *Improving teacher evaluation systems: Making the most of multiple measures*. New York, NY: Teachers College Press.

Haertel, E H (1991) New forms of teacher assessment. *Review of Research in Education, 17*(1), 3–29.

Hammonds, F, Mariano, G J, Ammons, G and Chambers, S (2017) Student evaluations of teaching: Improving teaching quality in higher education. *Perspectives: Policy and Practice in Higher Education, 21*(1), 26–33.

Hattie, J (2012) *Visible learning for teachers: Maximizing impact on learning*. London, England: Routledge.

Hayes, D (2014) *Factors influencing success in teaching English in state primary schools*. London, England: British Council.

Hayes, D, Chang, K and Imm, C (2011) Research and practice in language teacher evaluation. *English Teaching, 66*(1), 179–199.

Ho, A D and Kane, T J (2013). *The reliability of classroom observations by school personnel*. Available online at: www.k12education.gatesfoundation.org/resource/the-reliability-of-classroom-observations-by-school-personnel/

Hopkins, E, Hendry, H, Garrod, F, McClare, S, Pettit, D, Smith, L, … Temple, J (2016) Teachers' views of the impact of school evaluation and external inspection processes. *Improving Schools, 19*(1), 52–61.

Howard, A and Donaghue, H (eds) (2015) *Teacher evaluation in second language education*. London, England: Bloomsbury.

Hunt, B C (2009) *Teacher effectiveness: A review of the international literature and its relevance for improving education in Latin America* (Working paper no. 43). Washington, DC: PREAL.

Isoré, M (2009) *Teacher evaluation: Current practices in OECD countries and a literature review*. Paris, France: OECD Publishing.

Johnson, S (2013) On the reliability of high-stakes teacher assessment. *Research Papers in Education, 28*(1), 91–105.

Kane, T, Kerr, K A and Pianta, R C (eds) (2014) *Designing teacher evaluation systems: New guidance from the measures of effective teaching project*. San Francisco, CA: John Wiley & Sons.

Katsuno, M (2016) *Teacher evaluation policies and practices in Japan: How performativity works in schools*. London, England: Routledge.

Kaufman, D (2007) A multidisciplinary approach to assessment in teacher education, in Coombe, C, Al-Hamly, M, Davidson, P and Troudi, S (eds) *Evaluating teacher effectiveness in ESL/EFL contexts* (pp. 43–56). Ann Arbor, USA: University of Michigan Press.

Kim, K, Kim, G, Kim, S, Kim, J, Kim, J and Park, J (2010) *OECD review on evaluation and assessment frameworks for improving school outcomes: Country background report for Korea*. Seoul, Korea: Korean Educational Development Institute (KeDi).

King, M (2015) Evaluating experienced teachers, in Howard, A and Donaghue, H (eds) *Teacher evaluation in second language education* (pp. 167–179). London, England: Bloomsbury.

Kraft, M A and Gilmour, A F (2017) Revisiting the widget effect: Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher, 46*(5), 234–249.

Kuhlman, N and Knežević, B (undated) *The TESOL guidelines for developing EFL professional teaching standards.* Available online at: www.tesol.org/docs/default-source/papers-and-briefs/tesol-guidelines-for-developing-efl-professional-teaching-standards.pdf?sfvrsn=6

Lieberman, A, Campbell, C and Yashkina, A (2017) *Teacher learning and leadership of, by, and for teachers.* London, England: Routledge.

Liu, S and Teddlie, C (2005) A follow-up study on teacher evaluation in China: Historical analysis and latest trends. *Journal of Personnel Evaluation in Education, 18*(4), 253–272.

Liu, S and Zhao, D (2013) Teacher evaluation in China: Latest trends and future directions. *Educational Assessment, Evaluation and Accountability, 25*(3), 231–250.

Looney, J (2011) Developing high-quality teachers: Teacher evaluation for improvement. *European Journal of Education, 46*(4), 440–455.

Marzano, R J and Toth, M (2013) *Teacher evaluation that makes a difference: A new model for teacher growth and student achievement.* Alexandria, VI: ASCD.

Mateva, G, Vitanova, A and Tashevska, S (2011) *The European profiling grid: A user guide.* Available online at: www.epg-project.eu/wp-content/uploads/The-EPG-PDF-publication_EN.pdf

McCloskey, M L, Thornton, B and Touba, N (2007) New standards for a new era: Developing tools to assess the qualified teacher, in Coombe, C, Al-Hamly, M, Davidson, P and Troudi, S (eds) *Evaluating teacher effectiveness in ESL/EFL contexts* (pp. 7–24). Ann Arbor, USA: University of Michigan Press.

Miao, Z, Reynolds, D, Harris, A and Jones, M (2015) Comparing performance: A cross-national investigation into the teaching of mathematics in primary classrooms in England and China. *Asia Pacific Journal of Education, 35*(3), 392–403.

Mihaly, K and McCaffrey, D F (2014) Grade-level variation in observational measures of teacher effectiveness, in Kane, T, Kerr, K A and Pianta, R C (eds) *Designing teacher evaluation systems: New guidance from the measures of effective teaching project* (pp. 9–49). San Francisco, CA: John Wiley & Sons.

Ministry of Education, Ontario (2010) *Teacher performance appraisal: Technical requirements manual.* Ontario, Canada: Ministry of Education, Ontario.

Murdoch, G (2000) Introducing a teacher-supportive evaluation system. *ELT Journal, 54*(1), 54–64.

Murphey, T and Yaode, Q (2007) A China initiative: Portfolio-based teacher development, in Coombe, C, Al-Hamly, M, Davidson, P and Troudi, S (eds) *Evaluating teacher effectiveness in ESL/EFL contexts* (pp. 119–137). Ann Arbor, USA: University of Michigan Press.

National Board for Professional Teaching Standards (2010) *English as a new language standard* (2nd edn). Available online at: www.nbpts.org/wp-content/uploads/ECYA-ENL.pdf

National Board for Professional Teaching Standards (2016) *What teachers should know and be able to do.* Available online at: www.accomplishedteacher.org/wp-content/uploads/2016/12/NBPTS-What-Teachers-Should-Know-and-Be-Able-to-Do-.pdf

National Education Association (2010) *Teacher assessment and evaluation: The National Education Association's framework for transforming education systems to support effective teaching and improve student learning.* Washington, DC: National Education Association.

Nunan, D (2007) Standards-based approaches to the evaluation of ESL instruction, in Cummins, J and Davison, C (eds), *International handbook of English language teaching* (pp. 421–438). Boston, MA: Springer US.

Nusche, D, Laveault, D, MacBeath, J and Santiago, P (2012) *OECD reviews of evaluation and assessment in education: New Zealand 2011.* Paris, France: OECD.

O'Leary, M (2016a) Reclaiming lesson observation as a tool for teacher learning, in O'Leary, M (ed) *Reclaiming lesson observation: Supporting excellence in teacher learning* (pp. 1–9). London, England: Routledge.

O'Leary, M (2016b) *Reclaiming lesson observation: Supporting excellence in teacher learning.* London, England: Routledge.

OECD (2013a) *Synergies for better learning: An international perspective on evaluation and assessment.* Paris, France: OECD Publishing.

OECD (2013b) *Teachers for the 21st century: Using evaluation to improve teaching.* Paris, France: OECD Publishing.

OECD (2014a) *Talis 2013 results: An international perspective on teaching and learning.* Paris, France: OECD Publishing.

OECD (2014b) *A teacher's guide to Talis 2013: Teaching and learning international survey.* Paris, France: OECD Publishing.

OECD (2015) *Education at a glance 2015.* Paris, France: OECD Publishing.

Pennington, M C and Young, A L (1989) Approaches to faculty evaluation for ESL. *TESOL Quarterly, 23*(4), 619–646.

Peterson, K D, Wahlquist, C, Brown, J E and Mukhopadhyay, S (2003) Parent surveys for teacher evaluation. *Journal of Personnel Evaluation in Education, 17*(4), 317–330.

Pianta, R C and Hamre, B K (2016) Implementing rigorous observation of teachers: Synchronizing theory with systems for implementation and support, in Grissom, J A and Youngs, P (eds) *Improving teacher evaluation systems:*

*Making the most of multiple measures* (pp. 22–36). New York, NY: Teachers College Press.

Polikoff, M S (2015) The stability of observational and student survey measures of teaching effectiveness. *American Journal of Education, 121*(2), 183–212.

Purdon, A (2003) A national framework of CPD: Continuing professional development or continuing policy dominance? *Journal of Education Policy, 18*(4), 423–437.

Quirke, P (2007) A coherent approach to faculty appraisal, in Coombe, C, Al-Hamly, M, Davidson, P and Troudi, S (eds), *Evaluating teacher effectiveness in ESL/EFL contexts* (pp. 89–105). Ann Arbor, USA: University of Michigan Press.

Quirke, P (2015) A system for teacher evaluation, in Howard, A and Donaghue, H (eds) *Teacher evaluation in second language education* (pp. 101–114). London, England: Bloomsbury.

Raudenbush, S W and Jean, M (2014) To what extent do student perceptions of classroom quality predict teacher value added?, in Kane, T, Kerr, K A and Pianta, R C (eds) *Designing teacher evaluation systems: New guidance from the measures of effective teaching project* (pp. 170–202). San Francisco, CA: John Wiley & Sons.

Richards, J C and Farrell, T S C (2005) *Professional development for language teachers*. Cambridge, England: Cambridge University Press.

Santiago, P and Benavides, F (2009) *Teacher evaluation: A conceptual framework and examples of country practices*. Paris, France: OECD.

Santiago, P, Benavides, F, Danielson, C, Goe, L and Nusche, D (2013) *OECD reviews of evaluation and assessment in education: Teacher evaluation in Chile*. Paris, France: OECD.

Santiago, P, Donaldson, G, Looney, A and Nusche, D (2012) *OECD reviews of evaluation and assessment in education: Portugal 2012*. Paris, France: OECD.

Schultz, S E and Pecheone, R L (2014) Assessing quality in science teaching, in Kane, T, Kerr, K A and Pianta, R C (eds) *Designing teacher evaluation systems: New guidance from the measures of effective teaching project* (pp. 444–492). San Francisco, CA: John Wiley & Sons.

Seameo Innotech (2010) *Teaching competency standards in southeast Asian countries: Eleven country audit. Seameo Innotech*. Available online at: www.seameo-innotech.org/wp-content/uploads/2014/01/PolRes_TeachingCompetencyStandardsInSEA.pdf

Shakman, K, Riordan, J, Sanchez, M T, Cook, K D, Fournier, R and Brett, J (2012) *An examination of performance-based teacher evaluation systems in five states: Issues & answers.* Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands.

Shulman, L S (1988) A union of insufficiencies: Strategies for teacher assessment in a period of educational reform. *Educational Leadership, 46*(3), 36–41.

Speer, S and Harich, K (2005) *A description of responsibilities, structures and developments in the field of teacher and trainer evaluation*. Cologne, Germany: Univation, Institut für Evaluation.

Steele, J L, Hamilton, L S and Stecher, B M (2010) *Incorporating student performance measures into teacher evaluation systems.* Available online at: www.rand.org/pubs/technical_reports/TR917.html

Steinberg, M P and Garrett, R (2015) Classroom composition and measured teacher performance: What do teacher observation scores really measure? *Educational Evaluation and Policy Analysis, 38*(2), 293–317.

Stevens, R (2016) Observing what matters, in O'Leary, M (ed) *Reclaiming lesson observation: Supporting excellence in teacher learning* (pp. 141–151). London, England: Routledge.

Stewart, V (2013) *Teacher quality: The 2013 international summit on the teaching profession*. Asia Society. Available online at: www.asiasociety.org/files/teachingsummit2013.pdf

Stoynoff, S (2007) Building a context-specific teacher evaluation system for an ESL program, in Coombe, C, Al-Hamly, M, Davidson, P and Troudi, S (eds) *Evaluating teacher effectiveness in ESL/EFL contexts* (pp. 106–118). Ann Arbor, USA: University of Michigan Press.

Taut, S and Sun, Y (2014) The development and implementation of a national, standards-based, multi-method teacher performance assessment system in Chile. *Education Policy Analysis Archives, 22*(71). Available online at: www.epaa.asu.edu/ojs/article/view/1468/1313

The Centre for Development and Enterprise (2015) *Teacher evaluation in South African schools*. Johannesburg, South Africa: The Centre for Development and Enterprise.

Timperley, H (2011) *Realizing the power of professional learning.* Milton Keynes, England: Open University Press.

Treischl, E and Wolbring, T (2017) The causal effect of survey mode on students' evaluations of teaching: Empirical evidence from three field experiments. *Research in Higher Education, 58*(8), 904–921.

Tucker, P D and Stronge, J H (2005) *Linking teacher evaluation and student learning*. Alexandria, CA: ASCD.

Unesco-IICBA (2017) *Teacher support and motivation framework for Africa: Emerging patterns.* Available online at: www.unesdoc.unesco.org/images/0025/002599/259935e.pdf

Van de Grift, W J C M (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 J C M (2014) Measuring teaching quality in several European countries. *School Effectiveness and School Improvement, 25*(3), 295–311.

Van de Grift, W J C M, Chun, S, Maulana, R, Lee, O and Helms-Lorenz, M (2017) Measuring teaching quality and student engagement in South Korea and the Netherlands. *School Effectiveness and School Improvement, 28*(3), 337–349.

Wachtel, H K (1998) Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education, 23*(2), 191–212.

Wallace, T L, Kelcey, B and Ruzek, E (2016) What can student perception surveys tell us about teaching? Empirically testing the underlying structure of the Tripod student perception survey. *American Educational Research Journal, 53*(6), 1834–1868.

Weisberg, D, Sexton, S, Mulhern, J and Keeling, D (2009) *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness.* Washington, DC: New Teacher Project.

World Bank (2012) *What matters most in teacher policies? A framework for building a more effective teaching profession.* Available online at: www.siteresources.worldbank.org/ EDUCATION/Resources/278200-1290520949227/SABER-Teachers-Framework-Updated_June14.2012.pdf

World Bank (2018) *World development report 2018: Learning to realize education's promise*. Washington, DC: World Bank.

Worth, J, Sizmur, J, Walker, M, Bradshaw, S and Styles, B (2017) *Teacher observation*. London, England: Education Endowment Foundation.

Xerri, D and Campbell, C (2016) E-portfolios in teacher development: The better option? *ELT Journal, 70*(4), 392–400.

Youngs, P, Kim, J and Pippin, J (2015) Teachers' responses to changes in teacher evaluation policy in Korea and the United States. *International Perspectives on Education and Society, 27*, 413–442.

Zabaleta, F (2007) The use and misuse of student evaluations of teaching. *Teaching in Higher Education, 12*(1), 55–76.

Zhang, X-F, and Ng, H-M (2017) An effective model of teacher appraisal. *Educational Management Administration & Leadership, 45*(2), 196–218.

# Appendix: Using student learning in teacher evaluations

1    *The assessment of teachers' contributions to student learning should rely on multiple measures of student learning, not a single test or value-added score.*

2    *Measures of learning should reflect the curriculum a teacher is expected to teach and the range of skills and competencies students are expected to develop.*

3    *Valid measures should be used for all students.*

4    *Test measures intended to indicate growth must capture learning validly at the student's actual achievement level.*

5    *The use of any student learning measure should take into account factors that affect student achievement gains, including characteristics of the students and the context.*

6    *Value-added measures should only be used when there is a sufficient sample size and multiple years of data.*

7    *The evaluation system should consider evidence about student performance and teacher practice in an integrated fashion.*

8    *Various kinds of learning data should be considered in the evaluation process commensurate with their limitations.*

9    *The use of student learning evidence should be a source of continual study for educators, researchers and systems.*


(Source: Darling-Hammond, 2013, pp. 88–91)

This review assesses an international body of educational literature on teacher evaluation and draws on this to make recommendations for the evaluation of teachers in English language teaching. The report discusses components of effective teacher evaluation systems generally, followed by a detailed examination of a range of strategies which can be used in collecting information about how effective teachers are. Effective teacher evaluation emerges here as a multi-dimensional process that draws on several sources of robust evidence. It generates results that are used to support the professional development of teachers and, consequently, to improve student learning. The analysis also shows that the effective implementation of teacher evaluation can be hindered by a range of technical, financial, human resource, political, professional and social factors and that it is essential that these are given due attention when decisions about teacher evaluation are being made.

The British Council is the United Kingdom's international organisation for cultural relations and educational opportunities.