

Developing a self-assessment tool for English language teachers

Language Teaching Research

1–25

© The Author(s) 2018

DOI: 10.1177/1362168817752543
journals.sagepub.com/home/ltr**Simon Borg** 

Western Norway University of Applied Sciences, Norway

Adam Edmett

The British Council, Cultural and Education Section, China

Abstract

In the context of increasing global interest in teacher evaluation, this article describes the development of a self-assessment tool (SAT) for teachers of English and analyses responses to it by 1,716 teachers from around the world. One feature of respondents' self-assessments was that they were mostly positive, and this issue is discussed in relation to concerns about the accuracy of self-assessed competence more generally. Teachers also provided feedback on the tool itself and their comments on its relevance, clarity, coverage and value were mostly positive too. The teachers did, though, suggest ways in which the SAT could be developed further, and a number of design enhancements are discussed. Two particular challenges highlighted by the results of this study are also considered: the feasibility of developing a self-assessment tool that teachers of English in all contexts can use, and the extent to which teachers are able to assess their competence without reference to the specific circumstances in which they work. The potential for further research around the SAT is also noted, particularly in terms of how it can be combined with classroom observations to provide a more robust overall picture of what teachers are able to do.

Keywords

British Council, English language teaching, professional development, teacher evaluation, teacher self-assessment

I Introduction

Growing global interest in teacher quality has led in recent years to significant policy-related, empirical and practical activity in the field of teacher evaluation (for example,

Corresponding author:

Simon Borg, Klanec 21a, Kozina 6240, Slovenia.

Email: simon.borg@hvl.no

Avalos-Bevan, 2017; Bruns, De Gregorio, & Taut, 2016; Darling-Hammond, 2013; Kane, Kerr, & Pianta, 2014; OECD, 2013a). A recurrent point in this literature is that teacher evaluation should draw on multiple sources of information (Grissom & Youngs, 2016). One such source – teacher self-assessment – is the focus of this article. Specifically, we describe the development of a self-assessment tool, analyse results from its use with 1,716 teachers of English, and reflect critically on what these results imply for this specific tool and for the use of self-assessment in English language teaching (ELT), where research on teacher evaluation is not widespread (but see the papers in Coombe et al., 2007; Howard & Donaghue, 2015).

1 Teacher self-assessment

Self-assessment is a pervasive concept in education. The ability to assess one's knowledge, learning and performance is seen to be a key element in becoming an autonomous learner (Benson, 2011) while, in relation to professional learning, the importance of self-assessment has been stressed not just for teachers but also in other professions such as medicine (see Davis et al., 2006).

Various benefits of teacher self-assessment have been identified and in several state education systems around the world it is either encouraged (the General Teaching Council for Scotland, for example, offers teachers a 'self-evaluation wheel'¹) or a formal component of teacher evaluation (for example, in Chile and Portugal). One benefit of self-assessment is that it involves teachers more directly in teacher evaluation, giving them a greater sense of ownership in the evaluation process and in subsequent decisions about the areas of their work they need to improve (on the value of self-assessment in supporting professional development, see Borgmeier, Loman, & Hara, 2016; Ross & Bruce, 2007). It has also been argued that allocating some responsibility to teachers for the evaluation of their work is an appropriate way to recognize their status as professionals (see the discussion in Pennington & Young, 1989). Additionally, Marzano and Toth (2013) suggest that self-assessments (because they are informed by teachers' broad understandings of what they typically do) can provide a better picture of teacher competence than a small number of classroom observations conducted by an external evaluator. Self-assessment can take various forms (Bullard, 1998; MacBeath, 2003), from structured questionnaires to teacher portfolios (for the latter, see, for example, Alwan, 2007). In all cases, though, questions do arise about the accuracy of teacher assessments of their own competence, and we address these below.

While teacher self-assessment is not a new idea (Powell, 2000), ELT as a field globally (particularly in state education systems) remains characterized by top-down approaches to teacher evaluation where self-assessment is likely to be novel. However, in recent years a number of frameworks have emerged which can support the use of self-assessment in ELT teacher evaluation. These frameworks are underpinned by the view that, by reflecting in a systematic manner on what they know and can do, teachers can become more aware of the range of competences they need and identify appropriate directions for further development.

Developed for use in pre-service contexts, the European Portfolio for Student Teachers of Languages (EPOSTL²) includes a self-assessment section that contains '193

descriptors of competences related to language teaching ... These descriptors may be regarded as a set of core competences which language teachers should strive to attain' (Newby et al., 2007, p. 5). The British Association of Lecturers in English for Academic Purposes (BALEAP) also has a Competency Framework for Teachers of English for Academic Purposes³; various purposes for this are listed, including self-monitoring of professional development. Cambridge Assessment English has developed a framework to help teachers 'see where you are in your development – and think about where you want to go next.'⁴ Another self-assessment instrument is the European Profiling Grid⁵ which seeks to 'assist self-assessment and mapping of a range of current language teaching skills and competences' (Mateva, Vitanova, & Tashevskva, 2011, p. 12). The British Council has a self-assessment framework too; this is the focus of this article and we discuss it in detail below.

Despite the obvious current interest, though, in such frameworks, empirical activity focusing on their use by practising language teachers is scarce. Even in education more generally, while much has been written about teacher self-efficacy (for a recent review, see Zee & Koomen, 2016) and about approaches to professional development such as peer observation (Hamilton, 2013) and action research (Mills, 2014) which encourage teachers to evaluate their own teaching, limited work is available specifically on teacher self-assessment and on the use of competency frameworks to facilitate this process (some evidence from Chile, which we discuss later, is presented by Taut & Sun, 2014).

Outside education, one field where there has been substantial empirical work related to self-assessment is health and a particular concern in this research has been the extent to which health professionals can accurately self-assess their competence (for parallel analyses in the context of university students' self-assessment of their work, see Falchikov & Boud, 1989; Kearney, Perkins, & Kennedy-Clark, 2016). Based on their review of the literature comparing self-assessment with external measures, Davis et al. (2006) concluded that health professionals are limited in their ability to self-assess accurately, and this was often particularly true for less skilled individuals and for those who were most confident. Such conclusions about the validity of self-assessments echo those from earlier work, such as Mabe and West (1982), who, based on a review of studies where self-assessments were compared with performance, found a low correlation between the two measures. They also highlighted factors which improved the validity of self-assessments. Amongst these were the rater's previous experience with self-assessment and the anonymity of the self-assessor. Gaps between self-assessment and actual competence have also been noted by educational psychologists; Williams, Mercer & Ryan (2015, p. 45), for example, note that 'our sense of self may not necessarily be an accurate reflection of our actual abilities or performance.'

Overall, then, while, in the context of global contemporary interest in teacher evaluation, self-assessment by teachers is seen as one legitimate, formatively valuable though potentially imprecise source of evidence about teacher competence, limited empirical evidence is available of teacher self-assessment in ELT and this study responds to this gap. Such work is important not only to inform the further development of the specific self-assessment instrument we discuss here but to provide guidance for the design, use and evaluation of similar tools in the field of language teaching.

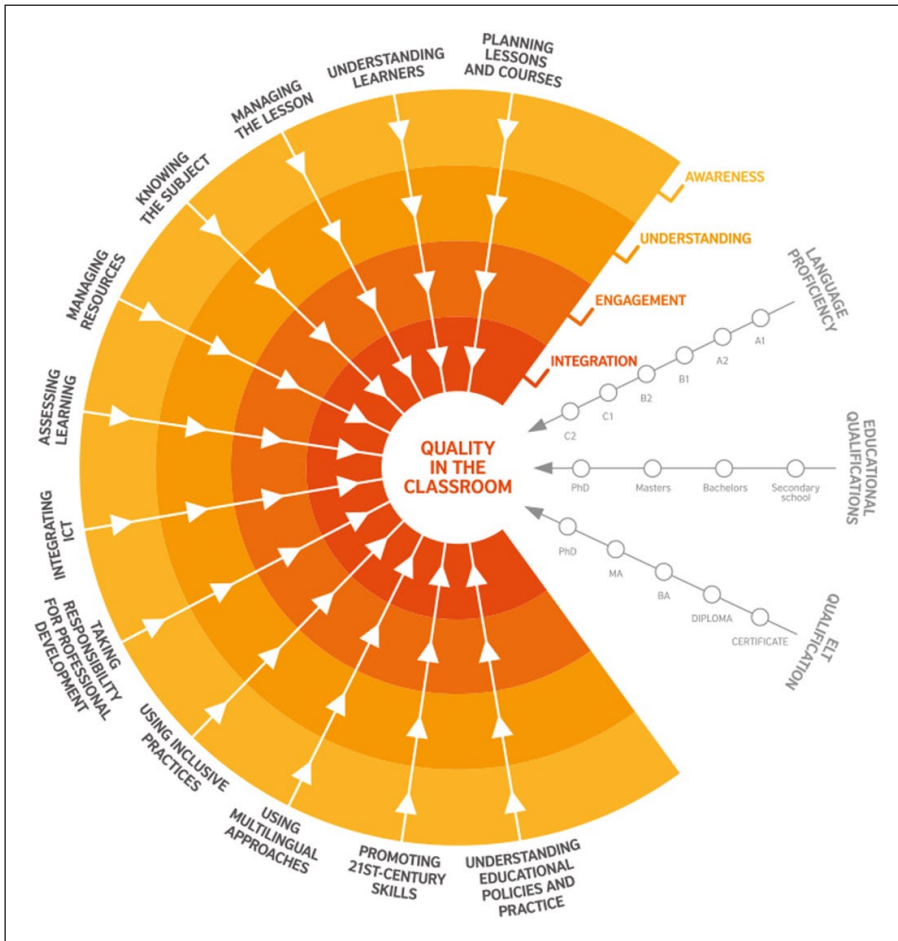


Figure 1. Global continuing professional development (CPD) framework.

2 Teaching for success

The tool we evaluate in this article is part of an approach to the professional development of English language teachers called ‘Teaching for Success’. Developed by the British Council, it includes, as Figure 1 shows, a global continuing professional development (CPD) framework (for an account of its origins, see Prince & Barrett, 2014) which has 12 professional practices, and each of these is broken down into more detailed ‘elements’ which describe what – according to the proposed framework – a teacher is required to know and do as part of that professional practice.

We elaborate on these elements below, but we would first like to explain the process through which the competences in the CPD framework (CPDF) were defined. There is, of course, no one universally accepted list of competences that teachers generally or English language teachers specifically need. The whole notion of competency

frameworks may even be rejected ideologically on the basis that teaching is too complex to reduce to lists of skills and knowledge that teachers require. Teaching is indeed a complex activity, but without some specification of target competences it is difficult to assess teacher quality and identify the professional development teachers need. Competency frameworks of the kind we are discussing thus seek to provide reference points against which such decisions can be made. Several examples from the field of language teaching were discussed above while competency frameworks from education more generally are also available; for example, that called ‘What teachers should know and be able to do’ (National Board for Professional Teaching Standards, 2016) has been very influential in the USA.

The professional practices in the CPDF were arrived at through a systematic process that involved, over some two years, feedback from external consultants, teacher development specialists, teacher trainers, teachers, Ministries of Education and other stakeholders. The process also involved an analysis of the contents of widely used ELT methodology texts (e.g. Harmer, 2007), the syllabi of internationally-recognized initial ELT qualifications (such as the CELTA course) and similar available frameworks (such as those noted above) to ensure that the content of the CPDF mirrored global understandings of core language teacher competences. Without denying alternative ways of conceptualizing teacher quality in ELT, the CPDF is thus the result of a systematic, formative and consultative process and provides an informed statement of what ELT practitioners are generally believed to need to know and be able to do.

3 *The self-assessment tool*

One of the instruments included in the CPDF is a self-assessment tool (SAT). The purpose of the SAT is to provide, against the professional practices in the CPDF, a measure of teacher competence which can be used by ELT practitioners globally and which (ideally in conjunction with other measures) can inform subsequent decisions about teacher professional development. In this article, our specific focus is on the design of the SAT and how teachers respond to it.

In their original form, the 12 professional practices in the CPDF were broken down into 139 individual elements describing what teachers know and can do. The SAT, though, needed to be an instrument that teachers could complete relatively quickly (30 minutes) and it was thus not feasible to include all 139 items. The final selection of content took place in two stages. First, three professional practices were omitted completely: ‘Using multilingual approaches’ was not felt to be globally relevant, while ‘Understanding educational policy and practice’ was very context-specific. A third professional practice, ‘Taking responsibility for professional development’, was already covered in an additional British Council needs analysis tool. Second, through consultation with ELT experts inside and external to the British Council, key elements for each of the remaining nine professional practices were chosen. The outcome of this exercise was a SAT with 48 elements, each phrased in terms of teacher ability or knowledge (see the Appendix). For most professional practices there were five elements, except for ‘Knowing the subject’ and ‘Promoting 21st century skills’ for which there were seven and six respectively (we review this and various other design issues later in the article).

It is important to note here that competency-based tools such as the SAT are also used as part of teacher evaluation in education more generally. Subject-specific examples exist, such as the self-evaluation of teacher effectiveness in physical education questionnaire (Kyrgiridis et al., 2014), as well as instruments which are designed for use across subjects (and which can be used for self-assessment). For example, Marzano and Toth (2013) present a tool that includes 41 teaching strategies and behaviours, while the widely used Framework for Teaching (Danielson Group, 2013) is organized around four domains, 22 components and 76 smaller elements. Teacher self-assessment using competency-based rating scales is thus not a novel idea; what is original, here, though, is the systematic study of this approach to teacher self-assessment in ELT.

Returning to the design of the SAT, one final decision related to the scale against which teachers would assess themselves, and this five-point scale was adopted:

1. The statement is not clear.
2. The statement is clear, but I'm not quite sure how to do this.
3. I can do this but not very effectively.
4. I can do this quite well.
5. I can do this very well.

This was used throughout the SAT except for the section on 'Understanding learners' which is about what a teacher knows about their learners (e.g. about learners' interests). In this case, the answer categories were:

1. I don't understand this question.
2. nothing at all.
3. a little.
4. a fair amount.
5. a lot.

In both cases, the first response option was included in this pilot to check the clarity of the statements on the tool; more broadly, though, it was felt to be important not to force respondents to assess their competence if they did not understand a statement, and this is another design issue we return to later.

The decision to use mainly closed questions was made with an awareness of both the benefits and drawbacks of such questions (for a discussion, see De Vaus, 2014); for example, while they are easier to complete and analyse, they also limit respondents' ability to express themselves fully. Overall, though, the literature on questionnaire design (e.g. Dörnyei & Taguchi, 2010) does recommend keeping open-questions to a minimum given the extra demands they make on respondents.

In addition to demographic items and the SAT statements, the instrument also included a final set of questions which invited respondents to comment on other aspects of the SAT, such as its length, content and value. A final open-ended question inviting any further comments was also included.

The draft SAT went through a series of reviews, including by external stakeholders. The resulting version was then piloted, and we report the results of this process below.

II Methodology

I Research questions

In establishing the context for this study, we highlighted the absence of research into teacher self-assessment in ELT and explained that our overall goal here was to critically evaluate how teachers respond to a specific self-assessment tool and to consider the implications of these responses for the development of this and self-assessment tools in ELT more generally. Accordingly, we address the following research questions:

1. How does a pilot group of English language teachers rate their competence on the nine professional practices on the SAT?
2. What are teachers' views about the SAT in terms of its value, relevance and content?
3. What suggestions do teachers make for improving the SAT?

2 Data collection and analysis

In partnership with the Open University, the British Council offers a course entitled 'Professional practices for English Language Teaching' on FutureLearn, which is a platform that offers MOOCs ('Massive Open Online Courses': free on-line courses). The course has been running since August 2015 and to date 164,644 participants have registered for it. In January 2016 and August 2016 all course participants were invited (via a link embedded in the course content) to complete the SAT. By clicking on this link respondents were taken to an on-line version of the SAT which had been prepared using SurveyMonkey.

Participation was voluntary and completing the SAT was not linked to any particular MOOC course requirement. When participants register with FutureLearn they agree that their data can be used for research purposes but, in line with the ethical guidelines of the British Educational Research Association (2011), respondents were also given additional information about how their answers to the SAT would be used together with guarantees of confidentiality and anonymity.

A total of 2,598 individuals started answering the SAT, though the dataset we examine here is smaller, as we explain below. Data were imported into SPSS 23 and analysis took the form of descriptive statistics. The final open-ended question (where respondents were asked for any other comments on the SAT) generated just over 3,500 words of text, which were analysed thematically using established procedures for coding and categorizing qualitative data (see, for example, Bryman, 2016). All responses were copied from SPSS into a Word document, read several times, and cut and pasted into different tables according to the themes they covered (e.g. benefits of the SAT or suggestions for improving it). Most answers had one central theme; a small number covered two themes and in these cases each theme was placed into its relative category. The analysis was initially completed by the first author, then a sample of data was checked by the second author; more specifically, the second author received 50 responses and a set of nine category headings and was asked to allocate each response to a category; we agreed on the analysis of 88% of these responses. The exercise was repeated with a slightly revised classification (with

eight categories) and agreement was reached in 92% of the cases. The remainder were agreed on through further discussion.

The ethical arrangements for participants in this study were explained above. It is also important that we are explicit about the roles of the co-authors of this article. The first author is not a British Council employee but has advised the organization on various projects, including the development of the CPDF. The second author works for the British Council's teacher development unit and has had key responsibilities for the development of the CPDF and the SAT. The motivation for writing this article is, though, the belief that it makes a contribution to the literature on teacher self-assessment in ELT rather than to promote the work of the British Council. While our focus in this article is on issues of design, we also note that the SAT is available for free use online.⁶ On completing it, respondents are presented with a visual summary of their results together with an automatically generated list of recommended British Council teacher development modules, some of which are free and others which can be purchased. How teachers use the results of the SAT, though certainly an interesting area for further study, is, however, not relevant to the analysis we present here and will not be discussed further.

III Results

As noted above, 2,598 teachers started the SAT. Of these, 372 said they were not teachers and they were removed from the analysis, together with a further 510 respondents who submitted largely incomplete SATs. The results we present here, then, are based on a non-random sample of 1,716 teachers of English.

I Profile

Of the 1,684 respondents who disclosed their gender, 84.6% were female. Table 1 shows that respondents worked in both state and private institutions, with slightly more in the latter category. Experience of teaching English varied (see Table 2), with the single largest group being the most experienced (over 15 years) and accounting for 29.3% of the sample. Respondents worked in 125 countries, though, as Table 3⁷ illustrates, almost 57% of the teachers worked in Europe (Ukraine, Russia, Spain, Italy and the UK were in descending order the five countries with most respondents). Almost 19% worked in Asia, with a similar figure for the Americas, but the remaining geographical areas were not well-represented. The teachers were also asked which age groups they taught (they could choose more than one), and secondary level ($n = 948$) was the most common, followed by post-secondary ($n = 773$), primary ($n = 607$) and kindergarten ($n = 158$). Two final questions in the introductory section of the SAT asked whether English was the teachers' mother tongue – 82.3% said it was not ($n = 773$) – and how they would describe their

Table 1. Type of institution.

Private	920	54.1
State/public	780	45.9
Total	1,700	100.0

Table 2. Experience teaching English as a foreign language.

	<i>n</i>	Percentage
0–2 years	238	13.9
3–5 years	316	18.4
6–10 years	362	21.1
11–15 years	297	17.3
More than 15 years	503	29.3
Total	1,716	100.0

Table 3. Respondents by geographic region (*n* = 1,708).

Region	<i>n</i>	Percentage
Europe	971	56.9
Asia	319	18.7
Africa	80	4.7
Oceania	17	1.0
Americas	321	18.8
Total	1,708	100.0

Table 4. How would you describe your own level of spoken English?.

	<i>n</i>	Percentage
Elementary	6	0.3
Pre-Intermediate	27	1.6
Intermediate	433	25.2
Advanced	1,250	72.8
Total	1,716	100.0

own level of spoken English. Table 4 shows that almost 73% described themselves as having advanced oral proficiency in English.

2 Professional practices

The core of the SAT consists of nine sections, each consisting of 5–7 items related to a particular professional practice. ‘The statement was not clear’ and ‘I don’t understand this question’ responses were filtered out and not included in the subsequent analysis of teachers’ self-assessments. Table 5 lists the 10 items that received the highest numbers of such responses. The item about ‘citizenship’ stands out here, but otherwise none were described as ‘unclear’ by more than 3% of the respondents. An analysis of these items does suggest cases where the wording could be clarified; e.g. a more specific explanation of ‘citizenship’ would be useful while the item on biases/beliefs is not linked directly enough to the professional practice on inclusive teaching that it is part of. However, in

Table 5. Items respondents found unclear ($n = 1,716$).

Item	<i>n</i>	Percentage
I can promote citizenship	159	9.3
I can use assessment results to inform subsequent teaching.	49	2.9
I can promote digital literacy	46	2.7
I can reflect on my own bias/beliefs and the impact this might have in the classroom.	42	2.5
I can locate appropriate digital content effectively.	38	2.2
I involve parents, learners and any other relevant persons in an inclusive learning environment.	31	1.8
I can promote student leadership and personal development	30	1.8
I can describe how learner understanding will be assessed.	28	1.6
I can describe how feedback on learner performance will be provided.	27	1.6
I can evaluate the quality of digital content.	25	1.5

Table 6. Reliability measures for each professional practice.

Professional practice	Items	<i>n</i>	Cronbach's alpha
Planning lessons and courses	5	1,653	0.78
Managing the lesson	5	1,704	0.78
Understanding learners	5	1,697	0.75
Knowing the subject	7	1,706	0.86
Managing resources	5	1,695	0.77
Assessing learning	5	1,653	0.81
Integrating information and communications technology (ICT)	5	1,664	0.89
Using inclusive practices	5	1,634	0.74
Promoting 21st-century skills	6	1,526	0.85

most cases there did not seem to be any obvious problems with the wording of the statements and any reported lack of clarity or understanding was most likely a reflection of teachers' unfamiliarity with the concepts being referred to, such as 'digital literacy'. It must be acknowledged, of course, that the number of 'unclear' responses reported here may be an underestimation; some respondents might have been reluctant to admit a lack of understanding while others may have unknowingly misunderstood statements.

Table 6 shows that all nine sections had internal reliability coefficients between 0.74 and 0.89 (0.7 is conventionally taken as an indication that items in a scale are addressing a common underlying construct; Howitt & Cramer, 2014).⁸ We can therefore sum up the individual elements in each section to generate a total section score and Table 7 and Figure 2 summarize teachers' overall self-assessments for each of the nine professional practices.

These results point to some overall trends in the teachers' SAT responses:

Table 7. Percentage frequency distribution of self-assessments.

Professional practice	n*	Not sure how to do this	I can do this but not effectively	I can do this quite well	I can do this very well
Planning lessons and courses	1,653	3.18	18.31	52.92	25.59
Managing the lesson	1,704	0.62	10.01	49.98	39.39
Understanding learners	1,697	3.87	11.65	38.76	45.72
Knowing the subject	1,706	2.57	22.26	51.87	23.30
Managing resources	1,695	2.83	13.64	48.22	35.30
Assessing learning	1,653	3.59	21.40	52.90	22.11
Integrating information and communications technology (ICT)	1,664	5.85	24.88	45.84	23.43
Using inclusive practices	1,634	2.49	17.40	47.80	32.32
Promoting 21st-century skills	1,526	8.19	25.70	48.00	18.11

Note. * n = Respondents who did not choose ‘The statement is not clear’ or ‘I don’t understand this question’ for any of the items in each professional practice.

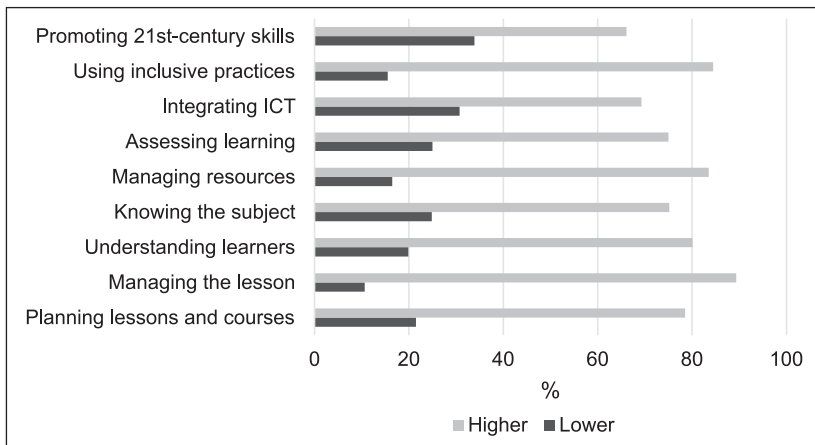


Figure 2. Summary of self-assessments on nine professional practices.

Notes. Higher = ‘quite well’ and ‘very well’; Lower = ‘not quite sure’ and ‘not very effectively’.

- Self-assessments across the nine professional practices were positive; when overall mean scores out of 4 on the 48 items are calculated for each teacher, for 44.9% the average rating was “I can do this quite well” and for 53.6% it was “I can do this very well”.
- This pattern was also evident across the individual professional practices, with teachers choosing ‘I can do this quite well’ or ‘I can do this very well’ in each case more than any other response.
- ‘Managing the lesson’ (89.4%), ‘Using inclusive practices’ (84.5%) and ‘Managing resources’ (83.5%) were the three practices with highest proportions of responses in the top two levels of self-assessment.

Table 8. Teachers' views about the self-assessment tool (SAT) (%).

	<i>n</i>	Agree strongly	Agree	Neither	Disagree	Disagree Strongly
The self-assessment was relevant to my context	1,701	33.5	52.5	11.1	2.5	0.5
The self-assessment was clear and easy to understand	1,703	41.3	49.5	7.7	1.2	0.2
Completing the self-assessment is a worthwhile activity	1,699	34.9	41.6	19.2	3.8	0.5
The self-assessment has encouraged me to look for further professional development opportunities	1,699	44.2	36.3	15.4	3.7	0.4
I would recommend the self-assessment to a friend	1,694	36.6	37.4	20.1	4.8	1.1
The self-assessment includes most of the skills, knowledge and behaviours that an English teacher needs	1,705	40.0	53.3	5.7	.9	0.2

- 'Promoting 21st-century skills' (33.9%), 'Integrating information and communications technology (ICT)' (30.7%) and 'Assessing learning' (25%) were the practices with the highest proportions of responses not in the higher two levels of self-assessment.

The Appendix lists all 48 items on the SAT according to the mean self-assessed competence on a four-point scale where 1=low self-assessment and 4=high self-assessment. This confirms the overall positive levels of teachers' self-assessed competence, with the lowest mean being 2.37. Reflecting the overall picture above, too, the lower part of the list is dominated by items related to 21st century skills (for example, critical thinking and problem solving or leadership and personal development), technology and assessment. One of the inclusivity items also appears low in the list – that referring to parental involvement – and we return to this item later in the article when we discuss the challenges that arise in designing an SAT that is relevant to a global audience of teachers of English.

3 Teachers' views about the SAT

The SAT also included a number of questions about users' views of it. Responses are summarized in Table 8 and these are overall very positive. Teachers thus felt that the SAT was relevant to their context, clear, and a worthwhile activity (although over 19% neither agreed nor disagreed that completing the SAT had been worthwhile). Teachers also agreed that the SAT encouraged them to look for professional development activities and that they would recommend the SAT to a friend (over 20% neither agreed nor disagreed here too). In terms of the content of the SAT, over 93% agreed or strongly agreed it covers most skills, knowledge and behaviours needed by teachers. One additional question

Table 9. Further Comments on the SAT ($n = 196$).

Focus of comments	<i>n</i>	Percentage
Generic positive remarks	47	23.7
Benefits of the self-assessment	39	19.7
Suggestions for improvement	38	19.2
Criticisms of the instrument	35	17.7
Other comments	16	8.1
Constraints on what teachers can do	14	7.1
Unclear comments	9	4.5
Total	198	100.0

asked respondents if they had completed the SAT within our target of 30 minutes. Over 91% said they had.

In interpreting these positive responses, it must be remembered that these teachers had voluntarily joined an ELT MOOC and were likely to be well-disposed towards professional development. Also, teachers who had not found the SAT a worthwhile exercise would have probably not completed it and thus the positive responses from those who did are perhaps not surprising; it would of course be interesting to understand why 510 respondents who started the SAT did not answer the self-assessment questions, but the anonymous nature of the responses did not allow us to explore this further. Overall, though, the teachers who completed the SAT expressed positive attitudes towards it.

4 Further comments

The final item on the SAT asked respondents if they wanted to make any further comments about it or their experience of completing it and 189 teachers did so, generating just over 3,500 words of open-ended comments. In most cases the comments addressed one distinct point but in nine cases they covered two, and thus the total number of discrete comments identified in the responses was 198. As a result of the thematic analysis described earlier, the categories listed in Table 9 were identified.

The largest category consisted of generic positive comments such as ‘excellent’, ‘great’, ‘thanks’ and ‘interesting’. Such comments did not specify a particular benefit of completing the SAT, but simply expressed a positive attitude towards it. Most comments consisted of a few words, though there were occasional longer contributions such as ‘I’ve done lots of pointless surveys in the past, but I think this genuinely is a useful task’ and ‘This is a very credible way of evaluating oneself’.

Moving on to the second largest category, almost 20% of the comments referred to benefits of completing the self-assessment. As the illustrative examples below show, promoting reflection on teaching and on areas for development and stimulating an awareness of what teachers need to be able to do were recurrent themes here:

- ‘This self-assessment shows areas that I need to improve.’
- ‘I found it really useful because it triggered some other questions about my practice.’
- ‘I think I have a better understanding of what is expected of me as a teacher.’
- ‘It made me think about what I was really doing in my classrooms and things that I should review.’
- ‘The self evaluation makes me discover my strengths/weaknesses and helps me develop them.’
- ‘This has given me the opportunity to reflect on my behaviour.’
- ‘This made me reflect on my teaching practice.’
- ‘I have learnt about key areas I should pay attention to as a teacher from this self-evaluation.’

A sizeable proportion of the further comments also either made suggestions for improving the SAT (19.2%) or were critical of aspects of it (17.7%; suggestions for improvement in such cases were implied rather than explicit). Over 63% of the suggestions ($n = 24$) related to the answer format used on the SAT, with various suggestions being made to allow respondents to express a fuller range of answers, as these examples show:

- ‘A couple of “I have never done this, but believe I would (not) be able to” options might be helpful.’
- ‘Change the possible answers to include “not relevant”.’
- ‘In the question about the amount between “a little” and “a lot” I would like to have a variant like “medium”/“some”/“enough”).’
- ‘Needs something between “... not very effectively” and “quite well”.’
- ‘The drop down list of options should have a “Not Applicable” option.’
- ‘I felt the choice of answers was too limited.’
- ‘Maybe there should be a “add a comment” section to each question. I would have liked to develop some more on a few questions.’

We return to this design issue later; for now it will suffice for us to state (as noted earlier) that we appreciate the manner in which pre-determined answers can limit respondents’ ability to express themselves but also understand the need for instruments of this kind to be easy and relatively quick to complete. Some respondents also, again rightly, noted that some of the demographic questions should have allowed a multiple rather than single response; e.g. regarding whether teachers work in state or private schools (it is of course possible to work in both).

Further suggestions made by the respondents identified additional topics that the SAT might cover (e.g. ‘I think that self-evaluation should include questions about time-management) and also recommended the inclusion of a demographic question on teacher qualifications which has since been added to the SAT. Finally, there were also suggestions about how the SAT might be followed up; e.g. ‘wonder if, based on the self-assessment, it could be possible for me to receive suggestions for particular online courses I might benefit from?’ or ‘I would have liked an instant rated summary of the self-evaluation’. Both these features have been incorporated into the latest version of the SAT.

In terms of explicit criticisms of the SAT, there were two particular themes: the first was that it was too long ($n = 9$) (e.g. ‘This is not a “short poll”,’ ‘It was a bit too long’ and ‘It was longer than I expected’). As noted earlier, over 91% of respondents said they completed the SAT within the 30 minutes they were told it would take, so perhaps this small number of complaints about the length was motivated more by the number of items on the instrument than by the time it took them to complete it. The second repeated criticism of the SAT ($n = 17$) was that some of the questions were not relevant to specific kinds of English language teaching. While, again, these comments came from a small proportion of the overall sample, they are important given that the ambitious goal of the SAT is to be a global tool for English language teachers. Respondents noted, though, that some of the questions were not relevant to teachers who taught individual students, very young learners, in adult education settings, or who taught specific skills only such as speaking. This perceived lack of relevance is interesting given that the SAT asks about what teachers can do rather than what teachers need to do in their current job, but the distinction is subtle and was in fact reflected in a further set of comments teachers made about factors that constrain what they can do, particularly in relation to the professional practice ‘Integrating ICT’, as in these examples:

- ‘I can use digital technologies but have no access to them in the classroom, so I wasn’t sure how to reply.’
- ‘Using digital materials ... In my school as in most schools in my country I don’t have an opportunity to use it very often and thus I can’t grow professionally in this area.’
- ‘In government rural schools where my practice is focused there is little done using ICT outside of specific technology classes.’
- ‘The place where I teach has no internet access nor availability of any electronic teaching aids, so that area of questioning, does not apply to me at present.’

The SAT invites teachers to self-assess their current competences (this is also made clear in the rubric at the start of the tool); in examples such as these, teachers were reflecting on their specific instructional contexts – on what they can physically do rather than on what they have the competence to do. Clarifying this distinction is an interesting challenge for the continuing development of the SAT and we return to it in the Discussion below.

Finally, there were a number of responses where the meaning was not clear (e.g. ‘Self-evaluation is a secret unseen thing inside me but it comes out and evaluates me when I do something good for others’) or which were beyond the scope of the question (e.g. ‘I want tips for being firm but friendly’). There were also a few isolated comments about the SAT that did not fit under any of the categories discussed earlier; one particularly thoughtful one, which resonates with our earlier discussion of the accuracy of self-assessments, was ‘I struggled between reality, honesty and modesty when completing it. There is a mess in my head.’

IV Discussion

In the context of increasing global interest in teacher quality and ways of evaluating it (OECD, 2013a, 2013b), teacher self-assessment is recognized as an option that can enhance teachers’ sense of agency and contribute formatively to their professional

development. While various frameworks for teacher self-assessment in ELT have emerged in recent years, critical analyses of their implementation and results remain scarce and this limits our understandings of their value. Our primary objective in this article has therefore been to evaluate a self-assessment tool for teachers of English. For this analysis, the tool was completed by 1,716 teachers from a range of countries (mainly in Europe, Asia and the Americas) and who were enrolled on an ELT MOOC. In this section of the article we will reflect on a number of issues raised by teachers' responses to the SAT. These issues are:

- trends in teacher self-assessments;
- accuracy of teacher self-assessments;
- design issues;
- developing a global tool.

1 Trends in teacher self-assessments

Across the professional practices in the SAT, teachers' overall assessments of their competence were high. Out of a total maximum score of 192 (48 items x 4, for the highest rating), the mean score for the teachers was 144.1, which is equivalent to an overall average rating of 'I can do this quite well'. The same trend is evident from the fact that, when responses for all professional practices are combined, 78% of the teachers said they could perform quite well or very well.

When individual professional practices are compared, the lowest rating was on 'Promoting 21st-century skills', where 64.4% said they could do it quite well or very well, followed by 'Integrating ICT' (69.3%) and 'Assessing Learning' (75%). Although the figures for these three practices are still positive, they do point to aspects of ELT where a proportion (in the range of 25% to 35.6%) of this sample felt less competent. 'Promoting 21st-century skills' (for a recent analysis, see Chu, Reynolds, Tavares, Notari, & Lee, 2017) is a topic that has achieved a raised profile in education more generally in recent years and it is not surprising to see this as the professional practice that teachers of English felt overall less competent in as there has not been much specific discussion of it within ELT. In terms of individual items (see the Appendix), four of the six elements for 21st-Century skills were in the five with the lowest ratings; the remaining two – creativity and imagination, and collaboration and communication – were higher up the list, which is also not surprising given their greater prominence in ELT.

At the other end the scale, the professional practices which were rated most highly by teachers were 'Managing the lesson' (89.4% said they can do this quite or very well), 'Using inclusive practices' (84.5%) and 'Managing resources' (85.5%). The first and third of these are very practical aspects of ELT which tend to feature quite prominently in training courses. 'Inclusive practices' is a less salient theme in the practical ELT literature – e.g. neither Harmer (2007) nor Scrivener (2011), two widely used ELT methodology texts, list 'inclusive' or 'diversity' in their index. However, it is not wholly surprising to find that teachers' self-assessments on the 'inclusive' items were high given the implications of, for example, suggesting that one did not 'treat all my learners equally and

with respect'. A closer analysis of these items does in fact suggest their design could be improved, and we return to this point in the relevant sub-section below.

2 Accuracy of teacher self-assessments

The second issue we focus on here, and one that was noted in the earlier discussion of the literature, relates to the extent to which teachers' responses to the SAT are an accurate reflection of their actual abilities. In the context of student self-assessment, this issue has been discussed at length (for a review, see Brown, Andrade, & Chen, 2015) and various reasons for inaccurate self-assessments have been noted; for example, individuals may have an inflated opinion of their actual ability or they may lack the information to self-assess appropriately. Both of these factors may have been at play here. With teachers, there will inevitably always be potential for inflated self-assessment to occur, given that (even under anonymous conditions such as those applied here) admitting limited competence may constitute for some individuals a threat to their professional identity (for a recent collection of papers on teacher identity in language teaching, see Barkhuizen, 2017). In other cases, teachers will genuinely believe they can do something quite well, even though this is not the case (this is more a case of limited self-awareness rather than an unwillingness to acknowledge the potential for improvement). Additionally, teachers' high self-assessments here were also likely a reflection of their profile as a self-selecting group voluntarily engaged in a professional development MOOC on ELT offered by a British university and therefore potentially already quite self-efficacious. We cannot, of course, dismiss the possibility that this sample just happened to be particularly competent, linguistically and methodologically and in support of this we can refer to results from smaller trials of the SAT where teachers' self-ratings have been less positive. For example, a group of 51 Korean teachers of English (49% of whom had more than 10 years' experience) recently completed the SAT and gave lower self-assessments on all nine professional practices compared to the group we have studied in this article; for example, the percentage of the Korean teachers who chose 'quite well' or 'very well' on planning lessons and courses was 53.7% (compared to 78.5% for the group studied here); the corresponding figure for assessing learning was 40.8% (75% here), for knowing the subject 40% (75.2% here) and for using ICT 43.1% (69.3% for our group here). Our understandings of trends in teachers' responses to the SAT will develop as further data from groups of teachers around the world becomes available. Most respondents in this study came from Europe and so more extensive trialling in other regions is necessary before firmer conclusions can be reached about the global utility of the SAT (we discuss the feasibility of designing a universal tool below).

With reference to students, various strategies for improving the accuracy of self-assessment have been recommended (see, for example, Ross, 2006). For instance, it is advised that a trusting learning environment can enhance self-assessment; in the context of teacher self-assessment this suggests that if teachers 'trust' that the exercise will have positive formative benefits for them they are more likely to provide a candid assessment of their competence (see below for further comments on formative teacher evaluation). Another suggestion is that self-assessment is more effective when students have time to make judgements with reference to specific pieces of work they have produced. Such

advice is difficult to apply in the context of a tool which is designed to provide a rapid assessment of teachers' competence in relation to a range of professional practices. While it would certainly enhance the process if prior to self-assessing teachers were able to see what competence in each professional practice looked like, this would make the administration of the SAT unfeasibly lengthy. One further suggestion that is relevant to the SAT, though, relates to the need for self-assessment rubrics to use language and concepts that are intelligible to users. This was an issue that was addressed when the SAT was being developed, and the feedback here from teachers on those items they did not understand (e.g. citizenship) will allow further refinements to be made in this respect.

One other factor that impacts on teacher self-assessments is the purpose of teacher evaluation, and in this respect the distinction between formative and summative teacher evaluation is very relevant (see, for example, Santiago & Benavides, 2009). While the former is concerned with improvement or development, the aim of the latter is to make a judgement related to, for example, contract renewal or promotion. In summative contexts the outcome of teacher evaluation has significant consequences for teachers and they will naturally want to inflate their assessments. Evidence of this was reported by Taut and Sun (2014), who analysed the use of self-assessments in a high-stakes teacher evaluation context in Chile and found that that score inflation was widespread (but unsurprising). They thus recommended that 'self-assessment should serve exclusively formative purposes' (p. 23). Teacher self-assessment is therefore more appropriate, and is likely to generate more accurate results, in the context of formative teacher evaluation, where the focus is on using the results to inform professional development rather than for accountability. Also, as noted by OECD (2013b, p. 18), 'just because the self-evaluation is not a valid evaluation for summative purposes, this does not mean it has no value. In fact, self-evaluation has great value in promoting professional development and teacher self-efficacy.' This is a key point about the SAT; it is conceived of as a formative tool rather than one designed to make summative judgements about teachers.

As already noted, too, assessments of teacher competence will be improved when multiple measures are used. For example, as Marzano and Toth (2013), suggest, self-assessments can be usefully combined with classroom observations. One practical challenge here, of course, is that while quantitative self-assessments can be completed rapidly by large numbers of teachers, classroom observations are labour-intensive and, to be reliable, need to be carried out by trained observers (see, for example, Ho & Kane, 2013 for a discussion of the challenges involved in obtaining reliable observational measurements of teaching). The British Council has developed a classroom needs analysis tool (CNAT) which focuses on the same professional practices covered by the SAT. As part of its trialling, the CNAT is being used by trained observers to describe the teaching of individuals who have completed the SAT. Early results indicate that the CNAT can moderate the SAT in a useful manner and further research into the relationship between SAT and CNAT results is planned.

3 Design issues

There are various aspects of the design of the SAT that can be improved. Some of these are minor, while others require deeper consideration. In terms of minor changes, a

question about teachers' qualifications has already been added to the latest version of the tool and it is important to avoid dichotomies where these are not appropriate (as noted earlier, teachers who work in both private and state schools should have the option to say so). The SAT would be more consistent, too, if every professional practice contained five items (as noted earlier, in one case there are six, and in another seven). Revising double-barrelled items (e.g. 'collaboration and communication' or those asking about both the range of techniques teachers can use and how engaging they are) to focus on one concept only would also improve the quality of the instrument. In two professional practices, too, the format of the questions varies; that on 'Understanding learners' is framed in terms of what teachers know, while that on 'Inclusive practices' is mostly about what teachers do rather than what they can do (we return to this distinction below). These are all design issues that can be remedied quite easily. Items which teachers said they found unclear can also be clarified, and in some cases this may require the avoidance of jargon such as 'digital literacy' or 'citizenship' which may have been problematic for some respondents.

A more challenging design question relates to the scale used on the SAT. In the version under discussion here, a five-point scale was used, with the first item on it giving teachers the option of saying that the question was not clear. It was felt to be important to provide such an option to assess the clarity of the statements. We are now aware of those which can be clarified and this will be reflected in the next major revision of the SAT. Moving forward, though, this item will be omitted, and where concerns exist that teachers may have difficulties understanding the questions the SAT will be translated into their language (as in a recent case in Peru). Some teachers also suggested in their comments that the answer scale be extended so that it provides more options. The literature which discusses the optimal number of response categories to include in questions of this kind (e.g. DeVellis, 2017) is not conclusive but it does stress that usability is one important issue to consider, and on that basis we do not think it is advisable to exceed five answer options, though the suggestion from our respondents that a 'not relevant' or 'not applicable' option be provided should also be considered and we will now reflect on that.

4 *Developing a global tool*

Readers will remember that the reason some teachers gave for wanting a 'not applicable' option was they felt that some items on the SAT were not relevant to their context (it is worth noting, though, that 86% of 1,701 teachers agree that the SAT was relevant). An obvious example is where teachers of adults were being asked about parental involvement. Less obvious is the item which asks teachers if they can supplement the coursebook – some teachers may not use one. Such issues arose because the SAT seeks to be useful to ELT practitioners anywhere. Based on teacher feedback here, though, questions arise about how feasible this is. Eliminating items (such as the one about parents) would not necessarily resolve the problem given the diverse range of contexts that ELT takes place and the many different forms and purposes that ELT can have. Perhaps the inclusion of a 'not applicable to my context' option at this stage would be useful to identify in further trials which particular items are marked in this way, and such information could

be used to further modify the items that are included under each professional practice. At the same time, though, the SAT asks teachers about their competence – what they can do rather than what they do do, and from this perspective including a ‘not relevant’ option is problematic given that the items included in the SAT are seen to reflect competences required by ELT professionals. Perhaps, though, this distinction between abstract competence and situated practice is too subtle; we appreciate, too, that it is only natural for teachers to self-assess their competence with reference to their experience, and accepting this may be the most productive way forward as the SAT is developed further. In fact, it could be argued that self-assessment which is grounded in what teachers actually do may be more accurate, just as asking teachers about their beliefs with reference to specific instructional episodes is likely to generate more realistic responses than when beliefs are elicited in a decontextualized or abstract manner (Borg, 2018).

Overall, then, based on the insights emerging here, we would propose that a revision of the SAT include five response categories, including one for ‘not relevant’ or ‘does not apply’ responses plus the same four-point scale of self-assessed ability that was used here. The most appropriate wording for the ‘not relevant’/‘does not apply’ option would need to be considered and trialled as teachers’ reasons for choosing this answer will vary; in some cases an item will simply not be relevant to teachers’ work (e.g. asking a teacher of adults about their interactions with parents) while in others the reason may be lack of opportunity (e.g. for teachers who want to use technology but lack access to it).

Even with these measures in place, though, we appreciate that the global nature of ELT and the many diverse geographical, institutional and personal contexts in which it occurs create significant challenges for the design of a universal teacher self-assessment tool. At this stage, though, we would not want to be unduly pessimistic; only 3% of respondents here felt that the SAT was not relevant to their context, while over 93% agreed that the SAT covered key teacher knowledge and skills (this finding suggests that the SAT’s focus on basic aspects of ELT, such as lesson planning, managing lessons and using resources is something that does enhance its global relevance). We are also encouraged by the fact that other ELT competency-frameworks we have mentioned also seek to be relevant to teachers of English generally and that, outside education, various frameworks exist that are seen to be widely applicable, sometimes even across different subjects. At this stage, then, and without in any way dismissing the challenges involved, we feel that there is value in developing the SAT further, based on the feedback emerging here, and to continue to assess its relevance in a range of global contexts.

V Conclusions

The increased availability of competency frameworks that support teacher self-assessment in ELT has created scope for critical analyses of their design, use, results and consequences that can contribute practically and empirically to both teacher evaluation and professional development. Work of this kind is currently limited, and in this respect the large-scale evaluation of a specific self-assessment tool that we have presented here constitutes an original contribution to the literature. Our analysis illustrates the value of extensive field testing when such tools are being developed; this is important to assess the content validity of the tool, to identify design issues that can be improved and to

highlight broader concerns about the quality of the results that merit closer scrutiny. In all three of these areas, the evaluation presented here has generated valuable insights that are relevant not just for the further development of this tool but for teacher self-assessment more generally.

Our analysis has identified various ways in which the SAT can be enhanced and highlighted challenges involved in developing a global tool. It also showed, though, that the vast majority of the teachers in this study felt that completing the self-assessment was a beneficial exercise. We did not, however, examine the consequences of these reported benefits and this is another interesting area for further research; for example, to what extent does completing a self-assessment result in concrete action by teachers to address areas of their work where they feel improvement is required?

Teacher self-assessment is, of course, just one of several options that are available to inform the broader process of teacher evaluation and self-rated competency checklists are in turn just one possible approach to self-assessment. We are not suggesting that the tool we have evaluated here is superior or preferable to alternative teacher evaluation strategies (and one point we have noted throughout is that teacher evaluation should be informed by multiple sources of evidence); our argument, though, has been that, although teacher self-assessment is valued in ELT and various frameworks exist to support it, this remains a largely unstudied area and much more systematic inquiry is needed into the development of such frameworks, how they are used, the results they generate, and the consequences of these results. We hope that this study will stimulate further examination of such issues.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Simon Borg  <https://orcid.org/0000-0003-2827-142X>

Notes

1. <http://www.gtcs.org.uk/professional-standards/self-evaluation/self-evaluation-standard-full-registration.aspx> (accessed January 2018).
2. http://archive.ecml.at/mtp2/fte/pdf/C3_Epostl_E.pdf (accessed January 2018).
3. <https://www.baleap.org/wp-content/uploads/2016/04/teap-competency-framework.pdf> (accessed January 2018).
4. <http://www.cambridgeenglish.org/teaching-english/cambridge-english-teaching-framework> (accessed January 2018).
5. <http://www.epg-project.eu/grid> (accessed January 2018).
6. <https://teachingenglish.english.britishcouncil.org> (accessed January 2018).
7. The United Nations classification was used here to group countries into geographical areas – see <http://unstats.un.org/unsd/methods/m49/m49regin.htm> (accessed January 2018).
8. We are aware of debates in the statistical literature about the use of Cronbach's alpha for assessing the internal reliability of scales (e.g. Sijtsma, 2009); as Pallant (2013), however, notes, it remains one of the most widely used measures of internal reliability, and our use of it here reflects common practice in social science research.

References

- Alwan, F. (2007). Teacher voice in teacher evaluation: Teaching portfolios in the United Arab Emirates. In: Coombe, C., Al-Hamly, M., Davidson, P., & Troudi, S. (Eds), *Evaluating teacher effectiveness in ESL/EFL contexts* (pp. 232–244). Ann Arbor, MI: University of Michigan Press.
- Avalos-Bevan, B. (2017). Teacher evaluation in Chile: Highlights and complexities in 13 years of experience. *Teachers and Teaching: Theory and Practice*. Electronic publication ahead of print version. Published online: 12 October 2017 (doi: 10.1080/13540602.2017.1388228).
- Barkhuizen, G. (2017). *Reflections on language teacher identity research*. London: Routledge.
- Benson, P. (2011). *Teaching and researching autonomy*. 2nd edition. Oxford: Routledge.
- Borg, S. (2018). Teachers' beliefs and classroom practices. In: Garrett, P., & Cots, J. (Eds), *The Routledge handbook of language awareness* (pp. 75–91). London: Routledge.
- Borgmeier, C., Loman, S.L., & Hara, M. (2016). Teacher self-assessment of evidence-based classroom practices: Preliminary findings across primary, intermediate and secondary level teachers. *Teacher Development*, 20, 40–56.
- British Educational Research Association. (2011). *Revised ethical guidelines for educational research*. Southwell: BERA.
- Brown, G.T.L., Andrade, H.L., & Chen, F. (2015). Accuracy in student self-assessment: Directions and cautions for research. *Assessment in Education: Principles, Policy & Practice*, 22, 444–457.
- Bruns, B., De Gregorio, S., & Taut, S. (2016). *Research on Improving Systems of Education (RISE): Measures of effective teaching in developing countries*. London: Department for International Development.
- Bryman, A. (2016). *Social research methods*. 5th edition. Oxford: Oxford University Press.
- Bullard, B. (1998). Teacher self-evaluation. Unpublished paper presented at the Annual Meeting of the Mid-South Educational Research Association, New Orleans, LA, USA. Available at: <http://files.eric.ed.gov/fulltext/ED428074.pdf> (accessed January 2018).
- Chu, S.K.W., Reynolds, R.B., Tavares, N.J., Notari, M., & Lee, C.W.Y. (2017). *Twenty-first century skills development through inquiry-based learning*. New York: Springer.
- Coombe, C., Al-Hamly, M., Davidson, P., & Troudi, S. (Eds.). (2007). *Evaluating teacher effectiveness in ESL/EFL contexts*. Ann Arbor, MI: University of Michigan Press.
- Danielson Group. (2013). *Framework for teaching*. Princeton, NJ: Danielson Group. Available at: <http://www.danielsongroup.org/framework> (accessed January 2018).
- Darling-Hammond, L. (2013). *Getting teacher evaluation right: What really matters for effectiveness and improvement*. New York: Teachers College.
- Davis, D.A., Mazmanian, P.E., Fordis, M., Van Harrison, R., Thorpe, K. E., & Perrier, L. (2006). Accuracy of physician self-assessment compared with observed measures of competence: A systematic review. *Journal of the American Medical Association*, 296, 1094–1102.
- De Vaus, D. (2014). *Surveys in social research*. 6th edition. London: Routledge.
- DeVellis, R.F. (2017). *Scale development: Theory and applications*. 4th edition. London: SAGE.
- Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research: Construction, administration and processing*. 2nd edition. New York: Routledge.
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, 59, 395–430.
- Grissom, J.A., & Youngs, P. (2016). *Improving teacher evaluation systems: Making the most of multiple measures*. New York: Teachers College Press.
- Hamilton, E.R. (2013). His ideas are in my head: Peer-to-peer teacher observations as professional development. *Professional Development in Education*, 39, 42–64.

- Harmer, J. (2007). *The practice of English language teaching*. 4th edition. Harlow: Pearson.
- Ho, A.D., & Kane, T.J. (2013). *The reliability of classroom observations by school personnel*. Seattle, WA: Bill and Melinda Gates Foundation. Available at: <http://k12education.gates-foundation.org/resource/the-reliability-of-classroom-observations-by-school-personnel> (accessed January 2018).
- Howard, A., & Donaghue, H. (Eds). (2015). *Teacher evaluation in second language education*. London: Bloomsbury.
- Howitt, D., & Cramer, D. (2014). *Introduction to SPSS in psychology*. 6th edition. London: Pearson.
- Kane, T., Kerr, K. A., & Pianta, R. C. (Eds.). (2014). *Designing teacher evaluation systems: New guidance from the measures of effective teaching project*. San Francisco, CA: John Wiley.
- Kearney, S., Perkins, T., & Kennedy-Clark, S. (2016). Using self- and peer-assessments for summative purposes: Analysing the relative validity of the AASL (authentic assessment for sustainable learning) model. *Assessment and Evaluation in Higher Education*, 41, 840–853.
- Kyrgiridis, P., Derri, V., Emmanouilidou, K., Chlapoutaki, E., & Kioumourtzoglou, E. (2014). Development of a questionnaire for self-evaluation of teacher effectiveness in physical education (seteq-pe). *Measurement in Physical Education and Exercise Science*, 18, 73–90.
- Mabe, P.A., & West, S.G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, 67, 280–296.
- MacBeath, J. (2003). Teacher self-evaluation. In: Keeves, J.P., Watanabe, R., Maclean, R., Renshaw, P.D., Power, C.N., Baker, R., Gopinathan, S., Kam, H.W., Cheng, Y.C., & Tuijnman, A.C. (Eds), *International handbook of educational research in the Asia-Pacific region: Part one* (pp. 767–780). Dordrecht: Springer Netherlands.
- Marzano, R. J., & 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., & Tashevska, S. (2011). *The European profiling grid: A user guide*. Available at: http://www.epg-project.eu/wp-content/uploads/The-EPG-PDF-publication_EN.pdf (accessed January 2018).
- Mills, G.E. (2014). *Action research: A guide for the teacher researcher*. New York: Pearson.
- National Board for Professional Teaching Standards. (2016). *What teachers should know and be able to do*. Available at: <http://accomplishedteacher.org/wp-content/uploads/2016/12/NBPTS-What-Teachers-Should-Know-and-Be-Able-to-Do-.pdf> (accessed January 2018).
- Newby, D., Allan, R., Anne-Brit Fenner, Jones, B., Komorowska, H., & Soghikyan, K. (2007). *European portfolio for student teachers of languages*. Available at: http://archive.ecml.at/mtp2/fte/pdf/C3_Epostl_E.pdf (accessed January 2018).
- OECD. (2013a). *Synergies for better learning: An international perspective on evaluation and assessment*. Paris: OECD Publishing.
- OECD. (2013b). *Teachers for the 21st century: Using evaluation to improve teaching*. Paris: OECD Publishing.
- Pallant, J. (2013). *SPSS survival manual: A step by step guide to data analysis using ibm SPSS*. 5th edition. Maidenhead: McGraw Hill.
- Pennington, M.C., & Young, A.L. (1989). Approaches to faculty evaluation for ESL. *TESOL Quarterly*, 23, 619–646.
- Powell, L.A. (2000). Realising the value of self-assessment: The influence of the business excellence model on teacher professionalism. *European Journal of Teacher Education*, 23, 37–48.
- Prince, E.S., & Barrett, A. (2014). Continuing professional development in action: An Indian experience. In: Hayes, D. (Ed.), *Innovations in the continuing professional development of English language teachers* (pp. 19–42). London: British Council.

- Ross, J.A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research & Evaluation, 11*, 1–13.
- Ross, J.A., & Bruce, C.D. (2007). Teacher self-assessment: A mechanism for facilitating professional growth. *Teaching and Teacher Education, 23*, 146–159.
- Santiago, P., & Benavides, F. (2009). Teacher evaluation: A conceptual framework and examples of country practices. Paris: OECD.
- Scrivener, J. (2011). *Learning teaching*. 3rd edition. Oxford: Macmillan.
- Sijsma, K. (2009). On the use, the misuse, and the very limited usefulness of cronbach's alpha. *Psychometrika, 74*, 107–120.
- Taut, S., & 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*. Available at: <http://epaa.asu.edu/ojs/article/view/1468/1313> (accessed January 2018).
- Williams, M., Mercer, S., & Ryan, S. (2015). *Exploring psychology in language learning and teaching*. Oxford: Oxford University Press.
- Zee, M., & Koomen, H.M. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being: A synthesis of 40 years of research. *Review of Educational Research, 86*, 981–1015.

Appendix I. Ranked means for 48 SAT items ($n = 1,716$).

Item	Mean
I treat all my learners equally and with respect.	3.68
I develop positive attitudes towards diversity in my classroom.	3.57
I understand my learners' level of English	3.45
I can create a positive learning environment.	3.42
I can give explanations that the learners are able to understand.	3.38
I can select materials from a range of different sources.	3.36
I can select activities which help meet the aims of the lesson.	3.27
I can give instructions effectively.	3.25
I can reflect on my own bias/beliefs and the impact this might have in the classroom.	3.23
I can select materials and resources based on learner needs.	3.22
I understand my learners' motivation to learn English	3.22
I can develop materials to supplement the coursebook I use.	3.20
I can check learners' understanding during the lesson.	3.19
I can give learners feedback on errors that helps them improve.	3.17
I can use a range of engaging techniques to teach vocabulary to my learners.	3.13
I can monitor learner engagement.	3.12
I can evaluate the effectiveness of the materials and resources I use during lessons.	3.12
I can write lesson aims which describe the intended learning outcomes for a class	3.09
I understand my learners' interests	3.09
I can use a range of engaging techniques to teach reading skills.	3.05
I can use a range of engaging techniques to introduce new grammar to my learners.	3.01

Appendix I. (Continued)

Item	Mean
I help my learners identify individual learning goals.	3.01
I can use a range of engaging techniques to teach my learners to speak English.	3.00
I promote collaboration and communication	3.00
I promote creativity and imagination	2.97
I can anticipate problems that may arise during the lessons and decide how to respond.	2.92
I can use a range of engaging techniques to teach listening skills.	2.91
I can use technology confidently for the purposes of teaching English.	2.91
I can measure learners' progress effectively.	2.90
I can assess learners in a range of ways.	2.88
I understand my learners' preferred ways of learning	2.87
I can use a range of engaging techniques to teach writing skills.	2.84
I can work with colleagues to design materials collaboratively.	2.84
I can describe how learner understanding will be assessed.	2.82
I can locate appropriate digital content effectively.	2.82
I can use digital tools effectively to help my students learn English.	2.82
I understand my learners' special educational needs	2.82
I can evaluate the quality of digital content.	2.81
I can describe how feedback on learner performance will be provided.	2.79
I can reflect on the effectiveness of the assessment I use.	2.79
I can use assessment results to inform subsequent teaching.	2.78
I can use technology to design and create teaching and learning materials.	2.78
I can use a range of engaging techniques to teach pronunciation to my learners.	2.73
I promote critical thinking and problem solving	2.66
I promote student leadership and personal development	2.66
I involve parents, learners and any other relevant persons in an inclusive learning environment.	2.64
I promote digital literacy	2.44
I promote citizenship	2.37

Note. 1 = low self-assessment; 4 = high self-assessment.