A partial replication of Thaine 2004, 'The assessment of second language teaching' in ELT Journal 58/4

by Helen Down

British Council's Master's Dissertation Awards 2020

Commendation

A partial replication of Thaine 2004, "The assessment of second language teaching" in *ELT Journal* 58/4

Dissertation submitted in partial fulfilment of the requirements for the award of Master of Arts in Applied Linguistics and English Language Teaching.

October 2019

Abstract

This study replicates one published in 2004 by Craig Thaine, which investigated how tutors on CELTA courses interpreted the assessment criteria for teaching which were in use at that time. The aim is to see whether changes introduced since the time of Thaine's study have affected the way the updated criteria for teaching are interpreted today. Following Thaine's procedure, the study draws on interviews with CELTA tutors, as well as a close examination of the literature on assessment in Second Language Teacher Education and the documentation accompanying the running of CELTA courses. The study yielded similar results to Thaine's, showing that there are criteria which are not consistently interpreted. Evidence also emerged that both CELTA tutors and Assessors themselves as well as those involved in the central administration of the qualification are aware of this inconsistency and of the need for a more substantial review than has been effectuated in over ten years. Based on the results of the study, some suggestions are made for how the assessment of the CELTA course might be improved.

Contents

7   Appendices

Appendix A   Full criteria used in the study, with descriptors

Appendix B   Overview of results

1   Introduction

1.1   The CELTA course

The course which is known today as CELTA (Certificate in Teaching English to Speakers of Other Languages), and which is recognised as a Level 5 qualification in the UK, has its origins in a programme developed by International House, London, in the 1960s and aimed at native speakers, usually of British English (Pulverness 2015: 11-12; Randall and Thornton 2001: 15). The idea was to provide them with the basic knowledge and techniques to teach English as a foreign language, without much formal background knowledge of grammar or linguistics, or of their future learners' L1 (Anderson 2018: 1; Hobbs 2013: 163-5). The organisation and management of the course have changed with time and there have been some updates to the content, but much of the original style and thinking behind the course remain, as does its usual format as a four-week, 120-hour intensive course (Hawkey and Milanovic 2013: 108; Poulter 2007: 2). The typical demographic of the course has also changed, with many candidates nowadays being non-native speakers and/or experienced (and in some cases qualified) teachers (Anderson 2018: 2; Wilson and Poulter 2015: 257; Hawkey and Milanovic 2013: 172). The most significant change, however, is the size of the undertaking. According to the official website, there are now 370 centres in 80 countries around the world at which the course can be taken (UCLES 2019a). CELTA today is administered by a body currently calling itself Cambridge Assessment English (UCLES 2019b), which is one of the examinations boards affiliated with Cambridge University. Most people associated with the qualification refer to this body simply as "Cambridge" and this convention will also be followed in this dissertation. Part of Cambridge's function is to try to ensure standardisation across the centres and courses around the world. They do this by having:

- two or three trained, approved tutors working on every course;
- each course moderated by an external, Cambridge-appointed Assessor[1] (many tutors are also Assessors);
- all tutors and Assessors complete a standardisation session once a year with results submitted to Cambridge;

---

[1] The term *Assessor*, capitalised, will hereafter refer to the official role of the external moderator of a CELTA course, while *assessor*, with a lower-case *a,* will refer to anyone who is assessing or evaluating teacher performance.

- assessed candidate work kept in a portfolio, stored by centres for six months after the course, available to Cambridge at any time;
- final grades agreed between course tutors and the Assessor, then submitted for approval by Cambridge. In the case of borderline pass/fail candidates and any others where there is any doubt about the recommended grade, the portfolios are sent to Cambridge for inspection;
- a set of assessment criteria, which both tutors and candidates have access to.

(Galaczi and Swabey 2015: 124 – 137; Cambridge English 2017)

The final point forms the focus of this dissertation. I chose this topic because I am a CELTA tutor and Assessor and have worked extensively within this course model, using these criteria on a day-to-day basis. While it is clear that they have been carefully and professionally produced, most tutors and Assessors, including myself, feel that there is room for improvement; they have also not been substantially updated in over ten years and few studies have been carried out into their use and effectiveness (see 3.1 below). To this end, I have chosen to examine the background to the development of the criteria and looked at how they are being used by tutors and Assessors today, with a view to establishing how consistent this is, what doubts and disagreements about them there are, and what the implications of these might be. The study is limited to criteria 1a – 5k, which are the ones directly concerned with classroom activity, for purely practical reasons. The criteria for planning and reflection, which have not been included, do of course interact and link with the ones discussed here, and most participants made explicit links to the former, though interestingly none to the latter during the actual discussion of the classroom criteria (a few forgot the instruction to stop at 5k and continued to talk about 5l – 5n). For the full criteria, with descriptors, see Appendix A.

1.2   Organisation of the dissertation

The dissertation consists of five chapters in total. Chapter 1 describes the background to the CELTA course and explains the choice of topic. In chapter 2, a critical review of key literature looks at the history of assessment in teacher education, with specific reference to Second Language Teacher Education, as well as at the principles behind assessment criteria and how they are drawn up. This includes exploring the background to the development of the CELTA assessment criteria themselves, as well as the link between modern teaching assessment criteria and the emergence of standards and competency frameworks. Chapter 3 presents the background and aims of the study itself, describes the method used to gather data and provides a profile of the participants who were involved. In Chapter 4, the results of the study are presented. This starts with an overview of the main approaches to criteria use which emerged, followed by analysis of participant interpretations of the individual criteria, organised in the order they appear in the CELTA assessment guidelines. Finally, Chapter 5 contains a discussion of the results and draws conclusions, relating the findings to Thaine 2004 and to the literature. The relative advantages and disadvantages of the current assessment criteria and the way

they are implemented are discussed and some suggestions are made for possible further study or amendment.

2   Literature Review

2.1   Assessment in Second Language Teacher Education (SLTE)

While it is clear that teachers have always been assessed or evaluated in some way, be it by superiors, colleagues, students or others, the process was rarely formalised in any sphere of education until the early twentieth century and did not start to gather momentum until the 1970s (Shinkfield and Stufflebeam 1995: 9 - 15). The history of assessment in SLTE, which in its current form is a young profession, existing only since the 1960s (Burns and Richards 2009: 2; Pulverness 2015: 11 -13; Howatt with Widdowson 2004: 245), has followed a similar trajectory to that of mainstream school education, with a similarly increasing awareness of a demand for professional accountability (Burns and Richards 2009: 6 -7; Malderez and Wedell 2007: 145) and a growing understanding of the need for reliability and validity of assessment (Thaine 2004; Galaczi and Swabey 2015: 114). As teacher assessment has progressed, there has also been a move towards trying to make it to some extent formative (to help teachers develop professionally) as well as summative (to give them a final mark or grade at the end of the assessment process), concepts which first arose in the 1960s (Shinkfield and Stufflebeam 1995: 22), and were further developed in the 1990s (Black and Wiliam 1998). Formative and summative assessment need not be mutually exclusive (Wallace 1991: 134), although on award-bearing courses this is difficult because the same tutors usually have to both *develop* and *assess* candidates, and many candidates are more immediately focused on their grade or result than on their overall development (Roberts 1998: 158; Freeman 1990: 116; Crooks 2004: 20). This general development in teacher education is also clearly evident in the assessment of SLTE (Thaine 2004: 336; Wallace 1991: 126; Galaczi and Swabey 2015: 123). In simple terms, assessment refers to, "making judgements about whether something is good or bad, acceptable or unacceptable, or whether something could be improved" (Rea-Dickins and Germaine 1992: 3). While this is something people do all the time in various areas of everyday life, one major reason formalised teacher assessment is a relatively new phenomenon is that the quality and effectiveness of teaching are, for a number of reasons, difficult to judge and the complexity of the challenge to do so has only relatively recently been appreciated. The main challenges are:

- Teaching is a complex process and 'good teaching' hard to define (Shinkfield and Stufflebeam 1995: 9; Roberts 1998:162; Wilson and Poulter 2015: 4; Graves 2009: 118; Korthagen 2017: 397 – 399; Freeman 1996: 353; Gebhard 1990: 157). Research into what constitutes good teaching has led, amongst other things, to attempts to pin down some of its components, such as use of different question types, time spent on tasks, 'wait time' before expecting learners to answer (Richards 1990: 5 – 13). These are *observable behaviours*, which a novice can, to a

certain extent, be taught. However, there are two problematic issues with observable teacher behaviours. Firstly, they are not merely techniques to be implemented randomly; for them to potentially have a positive effect on learning, a teacher needs to understand when and why they are used (Malderez and Wedell 2007: 12; Roberts 1998: 164; Hobbs 2013: 171). Secondly, it has gradually been realised that there is a lot more to teaching than observable behaviours. Wallace (1991: 6 – 7) talks of the *craft model* of education, in which novices are believed to be able to learn simply by observing, listening to and copying an expert in the field. This type of apprenticeship, described by Roberts as just one type of *model-based* learning (1998: 14 -18), originated in a behaviourist view of learning, was common in teacher education in the 1970s and 80s, and was indeed the method used to train CELTA tutors well into the 1990s (Morgan 2015: 152), meaning that many still working today were trained in this way. Most researchers do not completely reject the craft model (Wallace 1991: 16; Malderez and Wedell 2007: 15) but point out the dangers of relying too much on it in teacher education, as it is clearly at best only a part of the picture (Randall and Thornton 2001: 35 – 36; Darling-Hammond and Snyder 2000: 524).

- It is relatively easy to assess certain observable behaviours like those described above – or at least to notice that they have taken place – but some aspects of teaching and the effects of teaching on learners are harder to see, particularly in the course of one discrete lesson. This is part of the reason why courses like the CELTA include post-lesson reflection on the part of a trainee teacher prior to a feedback meeting at which they can analyse what happened in the class and present their view on it (Malderez and Wedell 2007: 152). The idea of *reflective practice,* which some writers feel has become a vague buzzword that many do not really understand (Burton 2009: 298), originated in the work of Donald Schön in the 1980s. It has been generally been very influential in training and education, though not without criticism (Roberts 1998: 50 - 53). Wallace suggests that, chronologically, the entirely practical *craft model* view of teaching described above was followed by what he calls the *applied science model* (see 2.2.1 below) and then by the *reflective model* (1991: 8 – 17). Although some writers think reflection is something novice teachers need to start learning to do as soon as possible (Randall and Thornton 2001: 161; Graves 2009: 118), the effectiveness of this type of reflection can be reduced on initial teacher education courses like the CELTA by the fact that beginner teachers lack the language to describe teaching (Roberts 1998: 34). Also they are conscious that their 'reflection' contributes towards summative assessment and thus anything they say may affect their grade (Delaney 2015: 95; Malderez and Wedell 2007: 162; Roberts 1998: 59 and 208). Delaney for one is very sceptical of the value of reflection in initial teacher education programmes and it should also be said that Schön himself had his doubts regarding its usefulness in this type of situation (Schön 1987: 326, 342; Roberts 1998: 53).

- Effective teachers find it difficult to identify and explain how they are effective because a lot of what they do has become instinctive through experience and practice. Research into *teacher cognition* has identified the importance of different types of decision teachers make both before and during teaching, but the experiential nature of many of these decisions makes it hard to teach people how to make them (Richards 1998: 10 – 12 and 29). Malderez and Wedell (2007: 25, 32, 34 – 38) explore the different types of knowledge teachers need in order to be effective: *knowing about* is conscious knowledge we can easily access, such as, for a language teacher, how to form the past tense; *knowing how* is connected with skills, which can at least in part be developed through observation; *knowing to* refers to realising that something should be done at a given moment, which then entails the other types of knowledge being employed. This latter is the type of knowledge that is gained mainly through experience. A large number of researchers nowadays adopt a broadly *constructivist* or *social constructivist* view of teacher learning, which argues that knowledge is constructed through the interaction of past experiences (including in this case the teacher's own schooling) with new input and experience (Wright and Bolitho 2007: 23; Roberts 1998: 44 – 46; Borg 2009: 164; Malderez and Wedell 2007: 14 – 15; Gebhard 2009: 254 – 255; Bailey at al. 1996: 11 – 17; Smith 1996: 207).

- Assessing teaching is at least partially subjective. Even when an assessor has a clear set of criteria to refer to, they will still have personal preferences in teaching styles, their own opinion of what constitutes effective fulfilment of the criteria, their own interpretation of learner responses, and days or moments when they are better able to concentrate than others. While it is generally accepted that some assessor variation is not only inevitable but desirable and that teacher assessment is never likely to become an exact science, there has to be some kind of control mechanism to ensure that the assessment serves the purpose it is intended to serve. It is for these reasons that assessors have to be carefully trained and that there needs to be a standardisation procedure in place (Alderson et al. 1995: 176; Malderez and Wedell 2007: 156 – 157; Roberts 1998: 157 – 160; Galaczi and Swabey 2015: 115, 124 – 135; Danielson 2012: 35 – 36; Gaies and Bowers 1990: 170 – 171; Strong et al. 2011: 368 - 369).

- Teaching is transitory. Although a lesson can of course be recorded and watched again, as is in fact often done for standardisation purposes (Galaczi and Swabey 2015: 132), most lessons are not recorded. Even when they are, the effects of the teaching in the moment it took place are over and cannot easily be recaptured or re-enacted, nor is the record likely to be complete or the behaviour of people who are being filmed the same as it otherwise is (Malderez and Wedell 2007: 162).

All this is true of most types of teaching, of course, but there are at least three further challenges which have been identified as specific to the field of SLTE. Firstly, there is a difficult and controversial relationship between theory and practice (Holliday 1998: 195 – 200; Richards 1990: 3 – 4 and 1998: 9;

Wallace 1991: 10 – 11; Ferguson and Donno 2003: 30). Both SLTE and applied linguistics, which can to a certain extent be seen respectively as the practice and theory sides of the same coin, date in their modern form from the 1960s (Burns and Richards 2009: 2; Richards 2008: 159). They have developed both together and apart, with substantial disagreement about the connection between them. While it is probably clear that someone could be interested and/or an expert in applied linguistics without being a language teacher or even in any practical way interested in language teaching, the amount and exact nature of the 'theory' a language teacher needs to know is more contested (Johnson 2009: 21 – 22; Ferguson and Donno 2003: 29 – 30; Lange 1990: 253; Ellis 2009: 136; Holliday 1998: 200 - 203). This links to the second difference peculiar to SLTE, which has been called "the dilemma of language as content" (Freeman et al. 2009: 79). For most subjects of study, there are two clear, though not mutually exclusive, areas for assessment: teacher *content knowledge*, which means how well they know their subject matter; and their knowledge of *methodology,* i.e. how to teach their subject. The meeting point or overlap between these two has been called *pedagogical content knowledge,* which describes knowing the main topics of the subject, what problems learners typically have with them, what preconceptions they have and how best to get key concepts across to them (Shulman 1986: 9). However, in the case of SLTE, teachers obviously need to know the language itself, know about the language, and know how to teach the language. This in turn links to the third difference, which is that second language teaching has both an established public sector in the form of ordinary state (and in some cases private) schools, where children learn languages alongside Maths and History and so on, and a 'private', supplementary, often less regulated sector, traditionally (and in many places still) the domain of the 'native speaker teacher' whose sole 'qualification' may be being a native speaker of the language being taught. The level to which teachers need to master the language itself, the depth in which they have to know about it, and even the methodology which is appropriate, can vary significantly according to the context in which they are working (Freeman et al. 2009: 81 - 87). This *situated* nature of second language teaching is something that is mentioned increasingly in the literature from the 1990s on (Darling-Hammond and Snyder 2000: 524; Johnson and Golombek 2018: 5; Roberts 1998: 163; Watkins et al. 2015: 308) and is a significant source of criticism of the type of SLTE represented by the CELTA and much of the 'private' language teaching sector (Ferguson and Donno 2003: 29 – 30; Anderson 2018: 11 – 12; Sulaimani and Elyas 2015: 25 - 26). In addition to these factors, Roberts (1998: 209) identified seven further difficulties he saw as particularly relevant to the CTEFLA course, which was the precursor to the CELTA. As these are all still relevant today, they are included here:

- pressure on tutors to pass candidates, and also provide a grade that enhances employment opportunities;
- little time (on intensive courses) between input and assessment, and the expectations that are fair to candidates in such conditions;
- candidates who are clearly improving but do not reach the pass standard after four weeks;

11

- candidates who start off well but show no change over the four weeks: can it be assumed that they will fail to develop after the course?
- candidates who meet practical criteria, but show little evidence of critical thinking about learners or teaching;
- candidates who do not teach very effectively but show good awareness of what they are doing and how it could be improved;
- the difficulty of defending differences between grades, given the aforementioned considerations and the fact that such forms of grading have known limitations.

Having established that SLTE is difficult to assess, it is now time to turn to how it is done. Assessment or evaluation[2] has an entire history of its own both inside and outside of teaching and has been studied in depth. According to Rea-Dickins and Germaine, in the context of education it needs to be, "systematic and undertaken according to certain guiding principles using carefully designed criteria" (1992: 2 – 3). It is clear from this that these researchers favour *criterion referencing* over *ipsative* or *norm referencing* for the assessment of teaching and this is also the method that has been chosen by Cambridge English, the body in charge of administering the CELTA. It is worth briefly looking at the difference. *Norm referencing* is the type of assessment where those being assessed are rated in comparison with each other, while *criterion referencing* compares them against a previously-agreed standard (Alderson et al. 1995: 156 – 157) and *ipsative assessment* compares them with their own previous performance (Bingham 2002: 9). Criterion referencing clearly has advantages if a qualification, like the CELTA, performs a 'gatekeeping' function, with those who pass being admitted into a profession (Malderez and Wedell 2007: 146 – 147). However, the creation of a clear, usable, reliable and valid set of assessment criteria is a substantial undertaking. Below are some considerations related to teaching in general, followed by a closer look at the thinking behind the assessment criteria used on the CELTA course.

2.2   Developing a set of assessment criteria for teaching

According to Dwyer, "[a]ssessment criteria judged to be technically, professionally, and legally defensible must proceed from an explicit conception of teaching and learning" (1995: 67). This is echoed by Richards with specific reference to SLTE (1998: 33). The choice and wording of the criteria

---

[2] While some sources differentiate between *assessment* and *evaluation* (see for example Nunan 1990: 62 or Malderez and Wedell 2007: 164), most choose one or the other, with *assessment* being the term used by Cambridge Assessment English.

will be dependent upon this conception, as well as on the aims of the course of study (Malderez and Wedell 2007: 150). Malderez and Wedell also recommend the use of *descriptors*, which are examples to illustrate an overall criterion and help make it more concrete; and they suggest the consideration of further descriptors to differentiate between different levels of achievement (2007: 152) (these are known on the CELTA as *performance descriptors* – see 2.2.2 and Cambridge English 2018b: 31). There is also the question of who should draw up the criteria, which then need to tested and revised, probably several times, before they are ready for use. Criteria need to be general enough to apply to a range of contexts yet precise enough for assessors to agree on what is meant and there should be neither too many nor too few of them (Malderez and Wedell 2007: 153; Shinkfield and Stufflebeam 1995: 74; Bingham 2002: 11; Roberts 1998: 171). Finally, in order for criteria to be used well, assessors need training in how to use them (Dwyer 1995: 72 – 79; Danielson 2012: 35 – 36; Galaczi and Swabey 2015: 129). It is also important to remember that, while a lot of assessment of teaching is done via direct classroom observation, which is largely what has been described so far, there are often several other components, which may include:

- lesson planning documentation;
- reflection on classroom performance (see 2.1 above);
- written assignments, for example in the form of essays to consolidate course work;
- a teaching portfolio containing some or all of the above;
- a reflective teaching journal;
- a final written exam or exams.

All but the last two are requirements of the CELTA, on which assessment is described as "continuous and integrated", where *continuous* means throughout the course and *integrated* that planning and teaching (which count as one assessment area) as well as the written assignments (essays) count towards the final grade (Cambridge English 2018a: 13 and 2017: 20). Planning documentation and written assignments obviously go some way towards mitigating the effects of the ephemeral nature of teaching and provide insight into candidates' thought processes and understanding. While reflective teaching journals have been shown to be quite effective in other situations (Richards and Farrell 2005: 68 – 83; Richards and Ho 1998: 153 – 170; Bailey 1990: 215 - 226) and are occasionally used in a casual way on some CELTA courses, they are time-consuming and probably seen by most providers as too much work on a short intensive course (Malderez and Wedell 2007: 155; Thornbury and Watkins 2007: 174).

2.2.1   Developing the CELTA assessment criteria

Full criterion-referenced assessment was introduced as part of a major revision in the mid-1990s, when the name CELTA was also first adopted (Pulverness 2015: 23; UCLES 1996 in Roberts 1998: 200; Thaine 2004: 338). There have been several updates since then, expanding from sixteen criteria

in 2000 to 41 in 2018, actually going down from 42 in 2017 (see 3.1 below). Comparing the guidelines above with the CELTA criteria, the first question to answer is, "What conception of teaching and learning is behind the course (and the criteria) and how explicit is it?". This is where Wallace's *applied science model* (see 2.1 above) comes in. Wallace describes it as the most common model behind most education programmes, which is based on practical skills being informed by scientific knowledge and empirical research (1991: 8 – 9). Richards (1998: 34 – 48 and 2002: 19 – 25) calls this approach the *science-research conception* of teaching. Although his distinctions and definitions are overall different from Wallace's (particularly his *art-craft conception*, which he uses to refer to a personal, contextual and experience-based way of teaching – in other words a constructivist view), the *science-research* and *applied science* ideas are broadly similar. Unlike Wallace (1991: 17), Richards believes it is not possible to mix the views, nor should any one of them be seen as 'right' to the exclusion of others. Instead, he suggests that they form a continuum, with the *science-research conception* being an appropriate starting point for new teachers, progressing ultimately to an *art-craft* approach, which presumably relies heavily on a substantial base of *knowing to*, although this is not a term he uses here.

Interestingly, in 1998 (: 48), Richards asserted that the RSA Certificate (predecessor to the CELTA) took an exclusively *science-research* approach, with candidates expected to produce language lessons following a strict 'presentation-practice-production' (PPP) format (Byrne 1986: 2 – 3; Anderson 2017: 218). This appears to be a criticism, although the Certificate was an initial training course and would therefore fit into his idea of *science-research* being the starting point. The comment is gone from an otherwise very similar chapter on the same topic four years later (Richards 2002: 22 – 24), which Thaine quotes when identifying the concept of teaching and learning which underpins the CELTA (2004: 337). It seems likely that Richards, like many people even more closely involved in the programme, fell into the trap of believing the Certificate set out to be more prescriptive than it did: as far back as 1977 it was noticed that, "the belief persisted amongst candidates, if not amongst tutors, that there was indeed a standard 'RSA lesson'" (Pulverness 2015: 17) – a belief which continues to afflict the CELTA to this day (Roberts 1998: 208), and with reason (see 5.2). Citing the same pages from Richards 2002, Thaine (2004: 337) initially asserts that the CELTA syllabus reflects a *science-research* conception of language teaching. Later (: 338), he says that "the syllabus is most probably based on a theoretical model (or models) of second language teaching" but that it is not the aim of his study to evaluate it. The current syllabus (updated since the one Thaine used but in this instance identical) does not explicitly espouse any particular conception of language teaching. The stated course aims are to enable candidates to:

- acquire essential subject knowledge and familiarity with the principles of effective teaching
- acquire a range of practical skills for teaching English to adult learners
- demonstrate their ability to apply their learning in a real teaching context.

(Cambridge English 2018a: 2)

Digging more deeply into the syllabus, there is evidence that observation of others is valued (: 3), that the importance of context is recognised (: 6), and that reflection is expected (: 11), all of which suggest a broad and probably mixed conception of teaching and learning. A slightly different impression emerges when examining the assessment criteria and, in particular, the descriptors, in more detail, where evidence of an espousal of Communicative Language Teaching (Ellis and Shintani 2014: 43 – 50; Howatt with Widdowson 2004: 326 – 350), as well as approval of a PPP approach is at least implicit in such wording as, "controlled oral practice activity", "communicative activity", "opportunity for communicative practice" (Cambridge English 2018b: 28). Roberts (1998: 210) offers what he describes as a "sensible 'constructivist' interpretation" of the CELTA, despite agreeing that providing the 'toolbox' of practical skills that is clearly one of the aims of the course goes against the idea of personalisation which is fundamental to constructivism. In his view, a novice teacher borrows the tools provided by the toolkit (in his words "dresses in borrowed clothes") for want of anything else and can personalise what they have learnt later. In order for this to work, he sees developing self-awareness (presumably through reflection) as essential. This interpretation is interestingly similar to that of Richards et al., who studied how a group of candidates processed the experience of learning to teach on the RSA Certificate course (1996: 258), and also reflected in Brandt (2008: 45).

2.2.2   The link to standards and competency frameworks

Galaczi and Swabey (2015: 117) state that, "[a]ssessment criteria [...] need to be explicitly based on an accepted theory or framework and to clearly and accurately conceptualise and operationalise that framework". Although they do not name any theory of teaching or learning as such, they refer to the burgeoning existence of 'frameworks' for 'profiling' teachers. Based initially on work done mostly with the US state school system and general education in mind, this has led to two frameworks directly designed to describe language teaching, the *EAQUALS European Profiling Grid* (2013) and the *Cambridge English Teaching Framework* (2014). In the introduction to the former (: 3), a constructivist, reflective view of language teacher development is described (though not labelled as such), while the latter was built on, amongst other things, the CELTA syllabus. The Cambridge *CELTA overview* (2015) links the CELTA to the "Foundation to Developing" framework stages of the Teaching Framework (although it is interesting and perhaps worrying to note that the Foundation stage describes teachers as able to provide, "accurate examples of language points taught at A1 and A2 levels", to which B1 is added at the Developing stage). These two frameworks were joined in 2016 by the *Cambridge English Trainer Framework*, which enables similar "profiling" of language teacher educators but is experience- rather than research based because of the current lack of research in this field (: 2; Wright 2009: 107). All of these initiatives and similar ones planned, for example, by Rossner (2009: 13) are strongly indicative of the growing trend towards what might be described as the *globalisation of standardisation*. The benefits are obviously increased transparency and faster recognition, for instance by employers and education providers, of what various levels and qualifications mean. However, there has been criticism of the "standards movement that currently

dominates education" (Graves 2009: 122), in particular that framework and competency statements are not created by or for individual institutions, are not easy to change, and that they force people to use descriptions that may not suit them or their context (Graves 2009: 122; Katz and Snow 2009: 74). These frameworks, which are very similar to each other and others like them, owe a lot to the work of Shulman, Grossman and, most recently Danielson (Galaczi and Swabey 2015: 117). Looking at an example of Danielson's (2014), which is designed for use in the US state school system, it can be seen that it is divided into four broad topic areas: *planning and preparation, the classroom environment, instruction* and *professional responsibilities.* Turning to the CELTA syllabus, we find five topics: *planning and resources for different teaching contexts, learners and teachers and the learning context, language analysis and awareness, language skill: reading, listening, speaking and writing, developing teaching skills and professionalism* (Cambridge English 2018a: 2)*.* It is clear from this that the first, second and fifth relate to general teaching skills, while the third and fourth are subject specific. The similarity to the Danielson framework is strong; even within the topic areas there are criteria and descriptors which are recognisably variations on the same thing.

The CELTA assessment criteria are thus linked to the five topic areas of the syllabus, which in turn can be related to broader areas of knowledge or competency. As seen in 2.2 above, it is important for assessment criteria to be neither to vague nor too precise, too many nor too few, in order to be valid, reliable and usable (Roberts 1998: 171 -173). Comparing again the Danielson framework, while the wording is very similar, Danielson has twenty-two criteria compared with the current forty-one on the CELTA. In addition to the criteria and the descriptors for each one, in 2013 a set of *performance descriptors* was introduced (Cambridge English 2018a: 13 – 14) to help with distinguishing between Pass, B- and A- grade candidates. Roberts points out that best practice would, in fact, be to eliminate B and A grades, given the difficulties described in 2.2 above, but that market pressures have led to them being maintained (1998: 209). One final point that is worth making here is that Cambridge do not specify how and to what extent they expect the assessment criteria to be used, which has led to variation and some questioning of their exact purpose (Douglas, 28.12.2018, personal correspondence; Bell English 2017).

3   The study

3.1   Background and aims

The purpose of the study was to partially replicate one carried out by Thaine in the early 2000s and published in 2004. Thaine was involved in the original design of the CELTA assessment criteria in the 1990s and was interested in following up this work to see how the criteria were being interpreted, with a view to suggesting amendments if any proved unclear or noticeably differently operationalised by tutors. In order to do this, he interviewed fourteen CELTA tutors working at various centres in New Zealand, asking them:

- What does each criterion mean to you?
- What trainee activity (-ies) in teaching practice do you look for in order to decide if the criterion has been met?

(Thaine 2004: 339)

He found that, while seven of the sixteen criteria were similarly interpreted, there was lack of clarity and overlap among the other nine. Feeling, as a result of his findings, that the criteria could be made clearer and thereby more reliable, Thaine suggested some amendments, which were adopted. Since Thaine's study was published, there have been several updates to both the syllabus and the criteria and the criteria have been extended, from sixteen in 2000 to 42 in 2017 and 41 in 2018, when two were merged. Since Thaine's study, it would appear that there have been few further studies of how the CELTA assessment criteria are used, and no replication. In 2013, Cambridge conducted a study which involved a survey sent out to 170 tutors and assessors in 40 different countries (Galaczi and Swabey 2015: 121). Although the survey itself is "internal and confidential" (Cambridge English, 25.10.2018, personal correspondence), the results were published by Galaczi and Swabey in 2015. The survey involved a series of statements to which tutors responded on a five-point scale from "very clear" to "unclear". Several of the statements took the form, "I have a clear understanding of the assessment criteria for…". Galaczi and Swabey conclude from the results that there is a high level of clarity within the criteria. Quoting the percentage of Pass, B and A grades worldwide over a period of four years, they suggest that the overall interpretation of the criteria must be sufficiently consistent. In 2017, Jacqueline Douglas, a CELTA tutor and Assessor who had noticed an increase in the frequency with which the criteria were referenced in lesson feedback (something which is not a requirement but also not prohibited), conducted an online survey of 54 tutors to find out whether they used the criteria in the feedback they gave candidates on their courses and their rationale behind doing or not doing it. She also questioned 12 CELTA course graduates to find out whether their tutors had used the criteria in this way and whether they considered it useful. She made her results known in a talk at the 2017

IATEFL conference (Douglas, 28.12.2018, personal correspondence). While Galaczi and Swabey's and Douglas's studies were obviously different both from Thaine's and from each other, two clear conclusions can be drawn from them which are of relevance to the current study:

- A large number of tutors are convinced they understand what the criteria mean;
- tutors are divided over how and when they use the criteria.

Douglas wrote to Cambridge to enquire about their position on criteria referencing and they said they did not mind as long as feedback was written in clear language, accessible to candidates. They stated that use of criteria is generally helpful to external Assessors and fellow tutors and that candidates require some support in interpreting them (Douglas, 28.12.2018, personal correspondence). While the main aim of the current study was to replicate Thaine's work, further evidence to support or contradict the findings of these other studies was also expected to emerge. This is relevant because, as seen in 2.2 above, research shows that in order to be effective, assessment criteria need to be sufficiently consistently interpreted and applied, as well as regularly reviewed (Shinkfield and Stufflebeam 1995: 60).

3.2   Method

Semi-structured interviews were conducted with thirty CELTA tutors and one tutor still in training. They were asked the same two broad questions Thaine had used and the same method was followed of not prompting them much, in order to obtain spontaneous data as far as possible; prompting was used to try to elicit clarification of ambiguous comments or to maintain focus, if needed. The tutors were encouraged to comment on the clarity and usefulness of the criteria and asked whether and how much they used the descriptors, which were added after Thaine's study with the idea that they would add clarity. In accordance with ethics requirements, they received an information sheet in advance and signed a consent form. The interviews were mostly carried out on a one-to-one basis, some face-to-face and some via Skype or similar applications. One focus group consisting of five participants, all colleagues at one centre but not all originally trained there, was also organised, and two participants were interviewed together at another centre, for practical reasons. The conversations were recorded as far as possible and the recordings transcribed.

3.2.1   Participant profile

Most of the tutors were working at the time of the interviews in centres around Europe. It is important to note, however, that many CELTA tutors are freelance workers who have travelled and worked in a number of countries. As a result, the experience they referred to in the interviews was based on

contexts ranging from the United States to the Middle East, former Soviet bloc countries, South East Asia, Australasia, and more. The least experienced tutor had shadowed one course and was not yet fully qualified, the most experienced had been working as a tutor for almost forty years. Those trained before the late 1990s underwent a less rigorous, non-standardised training procedure and the assessment criteria were introduced later in their careers, something that some of the older tutors commented on. It is a requirement of the training procedure nowadays that new tutors become familiar with the assessment criteria (Morgan 2015: 163). The current training procedure is described in Morgan 2015: 152 – 156 and Galaczi and Swabey 2015: 129 – 131. A final point to make is that, while the CELTA is the most widely-recognised qualification of its type available, it is not the only one and, particularly in the UK, there is some competition from the very similar Trinity CertTESOL (Trinity College London: 2019). Four of the tutors interviewed in the course of this study also had experience of working on Trinity courses.

4   Results

Thaine's main findings were:

- Seven criteria were consistently interpreted while nine showed variation
- The variations in interpretation were largely due to over-generalisation or overlap between criteria
- The variations could, to some extent, compromise the reliability of the criteria

(2004: 342 – 343)

It is worth noting that Thaine did not define "consistently interpreted" and, when asked, was unable to recall exactly how he had made this decision. He has, unfortunately, not kept his data (Thaine, 14.2.2019, personal correspondence).

Although the criteria were adapted and descriptors added in light of Thaine's findings, the current study yielded similar results. A high level of consistency was judged to be when 25 or more of the 31 tutors interviewed expressed a similar interpretation; reasonable consistency was 20 – 24, although it should also be noted that some tutors expressed mixed views or uncertainty, some skipped or conflated criteria, and in one case the first part of the interview failed to record, so these figures are only approximate. Using this interpretation, two of the twenty-four criteria investigated showed high consistency of interpretation, five reasonable consistency, and 17 fell below this. It should also be noted that several of the criteria which appeared to show a high or reasonable level of consistency in interpretation were operationalised in noticeably differing ways. This is examined in more detail below but to give one example, criterion 1d ("establishing good rapport with learners and ensuring they are fully involved in learning activities") showed 100% consistency in being interpreted as 'rapport', i.e. the relationship that the candidate builds with their learners. However, while *rapport* seems to be operationalised by most (around 24), explicitly or implicitly, as simply being friendly and pleasant towards the learners (and was therefore easily achieved, perhaps even to an "above standard" degree, by the vast majority of candidates), at least six or seven felt that superficial friendliness did not constitute good rapport and expected to see more response to learners, in some cases highlighting achievement of the fourth descriptor (learner involvement, which some even felt should be a separate criterion) and in one case laying great store by the second, which actually refers to time before and after class and which most tutors ignore. One tutor preferred to interpret 1d as "teacher presence", which for most others is part of 5d, although in fact the two are clearly linked in the syllabus (Cambridge English 2018a: 11). In some instances, there is a fine line between different *interpretation*

and different *operationalisation.* Appendix B contains a summary of the levels of consistency found, organised in the way Thaine did for ease of comparison. While Thaine analysed only a selection of his findings in detail, it is important in this more in-depth study to analyse all of them, as follows below.

4.1   Tutor approaches to criteria use

> One of the things that somebody I used to work for said to me is, "Make the criteria work for you. You know an A candidate, you know a B candidate; use the criteria to help you describe them, don't be criteria driven". So I use them as support for what I want to say, I think.
>
> [Participant in the current study]

As seen above, the literature on assessment criteria states that in order to be valid, reliable and useful, they need to be consistently and clearly interpreted. It is evident from the findings of this study that the CELTA assessment criteria are not entirely used in this way and, while there are what might perhaps be described as *trends* or *directions* in interpretation, or at least in operationalisation, these are sometimes wide-ranging. Some centres have made an institutional decision to interpret them in a particular way and issue guidelines to this effect; some individual tutors have made personal decisions of the same kind; in the case of one of the participants of this study, this is issued to candidates, although it is not always in line with the interpretation of his colleagues. There is even a website, which has nothing to do with Cambridge but which claims to "translate Cambridge Assessment English-speak into language comprehensible to normal people" (ELTConcourse 2018), by whom they mean CELTA candidates. Here all the criteria are "translated", in ways which do not always coincide with the findings of this study. In an interesting reply to an enquiry about this "translation", ELTConcourse (who remain elusively anonymous), said, "The aim was not to present some kind of paradigm concerning how the criteria should be interpreted (because that is unrealistic) but to present as nearly as possible what the language probably means to someone knowledgeable and experienced in the field" (ELTConcourse, 6.2.2019, personal correspondence). How important are such discrepancies? One noticeable trend in the findings of this study is that more experienced tutors have often decided what the criteria 'should' mean and pass these interpretations on to other tutors they train (their comments in the interviews sometimes even repeat each other verbatim); less experienced tutors are far more likely to raise the question, "What do Cambridge actually want here?", something Cambridge seem oddly reluctant to reveal. This study attempts to explore whether these differences matter, in other words to what extent and in what way they might affect the training and/or the results of the training that CELTA candidates are receiving.

4.2   Interpretations revealed in this study

Below is a description of what participants in the current study said, organised in the order of the assessment criteria for teaching valid at the time of the interviews (Cambridge Assessment English 2018b: 26 – 30).

<u>4.2.1   Topic 1 – Learners and teachers and the teaching and learning context</u>

It was widely felt that, in the words of one participant, this section is "a bit overladen". The majority of tutors interviewed shared the opinion succinctly summarised by one of them as, "Really, I merge all these together and I use 1d as the operator". 21 participants were largely dismissive of 1a ("teaching a class with an awareness of the needs and interests of the learner group"); 1b ("teaching a class with an awareness of learning preferences and cultural factors that may affect learning"); and 1c ("acknowledging, when necessary, learners' backgrounds and previous learning experiences"), or felt they could usefully be merged, as they see them as interdependent and covering similar ground; 12 suggested that they (or at least 1b) might be more relevant outside their current context, or where there is a larger cultural gap between candidates and students than is the case where most of them currently or usually work. Four participants expressed the view that this area was more in the tutors' hands than the candidates' because tutors usually know the students, they choose the materials, they pass on any relevant information to the candidates. Six felt it was more a planning than a teaching issue, to be included in "anticipated problems", which is planning criterion 4j ("anticipating potential difficulties with language, materials and learners"). Several felt it was unrealistic to expect candidates to show much knowledge in this area after two weeks (when the criteria are first mandatorily used) and that any real evidence of it was more likely to emerge in the *Focus on the Learner* written assignment (Cambridge English 2018a: 17) than in the classroom or, as five participants suggested, be a sign of a particularly strong candidate. Interestingly, around eight participants felt that, although Topic 1 mentions rapport and student needs, criteria that explicitly address the area of response to learners, the ability to listen to them, or what one participant called "reading the classroom" are in fact missing and that this tends to be put under 1d, 5d or 5h by default. One tutor, who is also responsible for recruiting teachers, said, "I think in real-life teaching 1d makes up 50%".

In summary, the overwhelming opinion of the tutors interviewed is that Topic 1 is too large and yet does not explicitly address the areas many want to (and generally do) interpret it as, namely a general response to student needs, in particular *emerging needs* during class. 1a – c are substantially ignored and 1d is usually used to cover the entire topic area. 1a, 1b and 1c all contain one descriptor which explicitly refers to planning rather than teaching ("use this information for selecting materials and activity types where appropriate"), for which no explanation is given.

<u>4.2.2   Topic 2 – Language analysis and awareness</u>

2a ("adjusting their own use of language in the classroom according to the learner group and the context") was described by everyone as "language grading", usually operationalised as amount and level of complexity, but sometimes tutors added clarity of articulation, speed of delivery, use of voice, and in one case accuracy, which for most comes under 2d. Seven tutors mentioned an overlap between this criterion, 5b and/or 5f, all of which refer to giving instructions, in the case of 2a and 5b in the descriptors. Some tutors were confused or bothered by this overlap, while others saw it as indicative of how important instructions are. Of those who feel that a criterion for "use of the voice" (projection, modulation, speed of delivery, etc.) is missing, some include it as part of this one, some with 1d, some with 5d and one with 5f, although the latter only refers to instructions.

2b ("identifying errors and sensitively correcting learners' oral and written language") is universally interpreted as "error correction", although there are some very different views on what is expected, broadly dividing tutors into three camps, one which feels immediate feedback should be prioritised, one which feels delayed feedback is all that can be expected of most CELTA candidates, and one which expects both. Here are typical comments from the three groups:

> I try to discourage them from doing too much post-activity correction
> because I think they find it very difficult […], they don't pick up on the right
> things.

> We encourage PACS [post-activity correction slot] for all our trainees and
> for the weaker ones it's enough.

> That trainees notice errors and that they are able to recognise at what stage
> of the lesson or what part of the stage they're at and then decide, based on
> where they're at in the lesson, what strategy they're going to use to deal with
> the error that has arisen.

There is a common perception, mentioned by around ten participants, that error correction is only to be expected towards the end of the course, although there is no mention of this in the syllabus or criteria; a minority of tutors (and one centre in particular) prioritise it from the outset. At least one tutor feels delayed error correction is often just a display of meaningless correction performed because the tutor expects it and another admits that candidates often 'plan' or make up the 'corrections' based on errors they might expect, rather than ones the learners have made. There is also some objection to the wording of the criterion, and in some cases its distance in the ordering of the criteria from its 'partner' criteria, 5h and 5j. The main objection to the wording is that it fails to explicitly address the concept of emergent language and responses to language gaps other than correction, such as feeding in missing lexis or providing more sophisticated alternatives. Some tutors include the latter under 5h, although it is not explicit there either. Finally, although correction of written language is mentioned in criterion 2b, only two tutors say it is something they make a point of doing; most others admit to ignoring it.

2c is "providing clear contexts and a communicative focus for language". Although one experienced tutor said, "This is one that is interpreted differently", half of those interviewed interpreted it almost exactly as he did, summed up quite neatly by one as, "a clear context from the beginning of the lesson, that they can exploit throughout the lesson, that they can keep referring back to. So, building from the beginning, through the whole lesson and using the context as much as possible". Several see it as more of a planning than a teaching issue, although a good number of those interviewed pointed out the tendency on the part of inexperienced teachers to fail to recognise and/or follow through on the context chosen (by themselves or their course book) at the planning stage. A couple of tutors felt that *context* and *communicative focus* should be separated, one of whom said he actually did so for his own purposes; three distinguished between a *topic* for the lesson as a whole and a *context* for language presentation; one tutor explicitly stated 2c is for him <u>not</u> a context for language ("I put that in 2e"), despite the wording of the criterion.

2d ("providing accurate and appropriate models of oral and written language in the classroom") shows noticeable inconsistency of interpretation, in two areas:

- Whether the criterion is relevant for native speakers, non-native or both;
- Whether the criterion refers to all language used by the teacher or to specific language items they have chosen to teach (called 'target language' by some).

Although one tutor felt that native speaker teachers are treated more leniently than non-natives with regard to this criterion, only five of those interviewed saw the criterion as aimed mainly at non-natives. Seven explicitly mentioned both and three only explicitly referred to native speakers, citing use of regional language varieties such as *youse* and 'incorrect' conditional models of the type *If I would have…*, as well as weak spelling. Interestingly, those who saw the criterion as aimed at non-native speaker teachers were more likely to be non-native speakers themselves (seven of the tutors interviewed were non-native speakers, three of whom referred to non-native candidates here).

It seems concerning that tutors are divided over whether this criterion refers to overall language use or only to language models explicitly being taught. As one of those in favour of the former said of his own language-learning experience, "What I remember is the classroom language, not the target language" – and it is obviously-salient classroom language that most tutors are quoting when they give examples of inaccurate non-native speaker language models: *how is it called?; very well* (as a form of praise); *what is your variant* [answer]*?; get in /beəz/* [pairs]. As another participant said, "This sort of stuff twenty times a lesson, you have to comment on it".

2e is "focusing on language items in the classroom by clarifying relevant aspects of meaning and form (including phonology) to an appropriate degree of depth" and is interpreted by all tutors as having to do with treatment of 'language' (usually meaning grammar, to a certain extent vocabulary and functional language items) in the classroom. Most differences in operationalisation which emerged

here seemed to be related to different ways of dealing with objections to both the wording and coverage of the criterion, which were mentioned by around 15 tutors. It is felt by some of these that the criterion is too densely-packed and they are often faced with the dilemma of what to do when they feel a candidate has met part of the criterion, perhaps even very well, and not met another part at all. Two tutors from the same centre said they write "part of 2e", while another mentions a centre which has gone its own way and, for internal purposes at least, split the criterion up, although obviously neither solution is ultimately satisfactory as for external assessment purposes the overall criterion needs to be graded to standard or not. A tutor who previously worked on Trinity courses feels that not having a separate criterion for phonology (which she points out is also missing from 3a with relation to listening) has led to candidates she sees now being "phonologically much less aware" than the ones she used to train. Most of the tutors who feel the criterion is overloaded advocate breaking it down into *meaning, form* and *phonology*, though two would like to see *grammar* and *lexis* separate. One tutor, without coming down on either side, reflected on the pros and cons of having so much together under this one criterion:

> […] they're not independent, sovereign items, they depend on each other. If you want to use this language, you need all three and from that point of view that's reflected in the fact that they make up one criteria. On the other hand […], if we're all breaking it up anyway […] it's obviously reflecting something we're seeing in the lesson […]. Counter-argument again, if you did separate it, are you giving grammar too much weight?

2f ("showing awareness of differences in register") is almost universally seen as an aberration, which is rarely referenced, considered met by default, and does not deserve to be a separate criterion. One tutor summed it up:

> I don't believe it should be a separate criterion at all as it's embedded and implied within 2e and the fact that it is separate also raises the problem of people misunderstanding or misconstruing what register is – as Cambridge I would say have themselves done in their glossary of it.

2g ("providing appropriate practice of language items") is described by all but two of the tutors who go into it in any detail (several consider it so self-explanatory they barely touch on it) in terms of a PPP lesson format (see 2.2.1 above), in which the 'practice' stage is taken to mean controlled practice (usually gapfill-type exercises) and 'production' freer practice (usually a speaking task, which some tutors use criterion 3b for – see below), believed by many to be necessary also in that order – which is how they interpret the third descriptor, "stage practice activities logically". The principal difference, of which they all seemed aware, is that some tutors insist on both 'practice' types being included and others feel that the omission of (usually) a classic 'freer practice stage' is something one can, in

the words of one tutor, "[F]orgive, as it were, sometimes not, depending where the problem came from – maybe they were being very responsive to problems learners had". One tutor perceptively observed that 2g does not in fact mention 'controlled' or 'freer' practice at all, but rather "appropriate" (the term "controlled practice" appears in none of the criteria themselves, just in two of the descriptors for 2b). Another suggests that, "[T]his distinction that's made between checking [understanding] and practising is a false distinction, it's a kind of continuum".

4.2.3   Topic 3 – Language skills – reading, listening, speaking and writing

Reading and listening are treated together under one criterion, speaking and writing under another. Speaking and writing were, until 2017, separate criteria. The criteria have also been reworded from *developing writing skills* to *practising writing skills.* Most tutors spoke about 3a ("helping learners to understand reading and listening texts") and 3b ("helping learners to produce oral and written language") together, so the same approach has been adopted here. Only three mentioned the change of wording, generally in a positive light, feeling that it was previously too ambitious. Around half the tutors were explicitly or implicitly happy with having receptive skills as one criterion and productive as another but some objections or questions were voiced, which included:

- Having them together might give the impression that the processes are the same in reading and listening (something at least 12 tutors actually seemed to believe), and in speaking and writing (although none of those interviewed appeared to believe this).
- Having them together might suggest that candidates only need to show an ability to teach one of each type, e.g. reading but not listening.
- Related to the previous point, and similar to the dilemma seen with criterion 2e above, how should the criterion be evaluated if a candidate shows themselves able to teach reading well but cannot teach listening or vice versa?
- Having speaking and writing merged leads to insufficient emphasis on writing. One tutor said, "I think it's largely down to tutors who've got a checklist of techniques which they want to tick off and writing lessons don't allow candidates to show those much". The argument that it was unrealistic or undesirable to have all candidates teaching writing lessons was explicitly voiced by five. One tutor admitted he found writing lessons boring and so was happy not to have to include them. Having writing and speaking together is, in fact, commonly interpreted as 'permission' to exclude writing, although no one seems to think excluding reading or listening, let alone speaking, would be acceptable.

One participant commented that it would make "just as much sense" to have reading and writing as one criterion and speaking and listening as another. In fact, on CELTA courses, candidates are generally trained to teach reading and listening skills using a similar basic procedure, which several tutors referred to as "the recipe". Most tutors did not question this standard procedure (and most

course books also follow it); of the six who did, five seemed to think it was worth introducing other methods and three mentioned the absence of a link between phonology and listening; most, however, appeared to feel anything other than "the recipe" would be beyond novice teachers or too much for the scope of the course. Nine tutors mentioned that Topic 3 lacked detail; two suggested that more, or clearer, mention of "sub-skills" could be included in 3a and another thought the specific mention of *top-down* and *bottom-up processes* would be useful. It also emerged that at least eight of the tutors use 3b not in the context of speaking lessons as such but rather for the 'freer practice' stage of a grammar lesson or the follow-up to reading or listening. The loosest interpretation of 3b, voiced by one and perhaps implicit in some others, was, "Showing that they're able to let learners speak as much as possible". One tutor appeared to think that 3b might be intended to be linked to 3a: "Should it be like, 'Produce oral and written language related to the text'? Or to the topic of the text?"

To summarise Topic 3, concerns are raised both about having clearly very different skills put together under the same criterion and about the effects this has on the way they are treated on the course. There is some feeling that there is a lack of detail here and an imbalance compared with Topic 1 and/or 2.

4.2.4   Topic 5 – Developing teaching skills and professionalism

5a ("arranging the physical features of the classroom appropriately for teaching and learning, bearing in mind safety regulations of the institution") was the one criterion that provoked immediate laughter from a number of tutors and from one the comment that he could not imagine it being graded *below standard*, "unless you blocked the doorway and struck a match". While clearly facetious, his point was that the criteria are designed to be potentially graded *at standard, above standard* or *below standard* and yet for some he felt a *below* or *above* standard rating made little sense, something another tutor also mentioned here. Most tutors claimed rarely to comment on 5a, usually because they are not able or permitted to make changes to their classroom layouts but also because they considered the criterion generally unimportant. A few said they liked to see candidates experimenting with different layouts but did not appear to insist on it. Only one of the tutors interviewed enthused, "5a is one of my favourites! This is my hobbyhorse because I'm just not a fan of horseshoe classrooms". And one other said he had finally "found a use for it" – to refer to the teacher's position in the classroom, which is also one other tutor's interpretation of it. In summary, the criterion shows a high level of consistency of interpretation as something like *varying the layout of the room to suit different group sizes and activities*, with lip-service to safety regulations some tutors claimed not even to know themselves. This is similar to one of the main interpretations of 5b (see below) and one tutor suggested 5a and 5b could perhaps be merged.

5b ("setting up and managing whole class and/or group or individual activities as appropriate") was one of the least consistently-interpreted criteria and one which provoked considerable confusion. It was also one of the ones for which some of the more experienced tutors had decided on a personal interpretation. For two tutors, it referred to allowing learners to check answers in pairs (which for most

fell under 5h, where it is an explicit descriptor); for four it referred to the choice of interaction pattern for an activity and in a similar vein two tutors felt it was phrased more like a planning criterion, for choosing interaction patterns in advance (which is planning criterion 4f: "including interaction patterns appropriate for the materials and activities used in the lesson"). A similar interpretation was actually organising learners into the chosen pattern, perhaps carrying through with or adjusting from the one planned. Overlap with 5f and 2a was mentioned by several, one or two of whom appeared only to notice it during the interview: "Haven't we just…isn't that repeated? I think they [5b and 5f] could be merged, they seem to be repetitive". An interesting comment on the overlap was, "I would find it more logical if *setting up* were together with *instruction giving*, so if the first part of 5b were with or in 5f because *setting up* and *managing* activities are very different things". Some felt that, while the criteria themselves were worded differently (though it seems not transparently, as one tutor who refused to refer to the descriptors during the interview felt unable to comment on 5b at all), the presence of one similar and one identical descriptor muddied the waters. Three tutors thought it could usefully be put closer in the list to 5f, though not necessarily merged with it. The most idiosyncratic interpretation was, "To me this is just a matter of having classes where the teacher hardly seems to be present and it's just students doing activity after activity. Or it's not the teacher that's dominating the classroom […]".

5c ("selecting appropriate teaching techniques in relation to the content of the lesson") was not so much inconsistently interpreted as self-confessedly not understood by five tutors, while some others (unusually and therefore noticeably) said, "I <u>think</u> it means...". Adjectives chosen to describe it included *nebulous*, *vague, broad, general* and *wishy-washy*. One tutor said, "It's a whopping great thing to just have one [descriptor]", which perhaps explains why several tutors found it hard to understand. Of those who did feel they understood it, around six felt it to be more a planning issue, though two made the point that it might show an understanding of planning in the way it was carried through in class. The most consistent interpretation, shared by around fifteen tutors, was that it referred to choosing suitable staging for the type of lesson being taught and, within that, executing the stages in ways that made sense (as one put it, "Not drilling the true/false statements"). There was some feeling that this area was covered elsewhere ("For skills it's in 3a and for language it's sort of in 2e really. So I think you could get rid of [5c]"; "It's basically repeating 3a, 3b"). Two tutors raised more serious questions about it, one saying, "It does imply that there are certain techniques that are appropriate for certain types of lesson – it would be useful to know what Cambridge thought that was", and another, "I don't feel the restrictions or divisions or categories and adherence to them is something that we need to do or is beneficial".

5d ("managing the learning process in such a way that the lesson aims are achieved") was another criterion which showed a particularly low level of consistency in interpretation, confessions of lack of understanding and it was noticeable that different tutors tended to prioritise different descriptors. Several of them were clearly aware of these differences, (it, "seems to have so many different

interpretations", "[is] open to interpretation", "[has] serious over-reach and ambiguity"). Two tutors mentioned it as "a catch-all", which one described as being used by many "to sort of justify why a lesson didn't function very well". The word "nebulous" appeared again twice, someone described it as a "vague one, for me at least […] it encompasses so much" and another said, "I find that description 'managing the learning process' a bit sort of airy-fairy, I don't find it helpful, I don't know what that means". The main interpretations were as follows:

- Eight tutors explicitly said it was mainly about "achieving aims", though doubts were raised about the actual operationalisation and even importance of this, one observing that it can be hard to judge in forty minutes or so whether a learning aim has been achieved, and two that the aims may be ill-chosen in the first place. Related to "achieving aims" were comments about the criterion being, "so people don't go off on tangents" or "so there's not just any random activities".

- The other main interpretation prioritised the other four descriptors over the "aims achievement" aspect, with two tutors stating that it was all about, "changing your role in the classroom", for instance being a facilitator or more controlling, as required. Two emphasised the importance of knowing when to be directive, firm or in the words of one "teacher-like". One explicitly extended this idea to a response to learner needs.

- It did indeed appear to be a "catch-all" for at least two or three tutors, the most explicit expression of which was, "This is where I think, 'there's something wrong with this lesson'. I use this – this is terrible actually – I use this to fail candidates." In the same vein, two tutors felt it was only ever mentioned as a weakness, three claimed to use it for failing candidates and, related to this, one pointed out that it often "crops up in warning letters" (a required warning letter is written towards the end of the course if a candidate is at serious risk of failing).

Further comments which came up several times were that it is related to pace and timing (five tutors) and that it is related to monitoring (six tutors). That is was fundamentally important came out in different ways including, "it undercuts everything else" and "it's what we're supposed to be looking at, really". Only one tutor felt it could be subsumed elsewhere.

In summary, while everyone seemed to at least sense that this criterion is important, there were marked differences in what tutors expected to see in order to consider it met. While "achieving aims" and "being directive when required" probably sum up the two main interpretations accurately, few tutors appeared to see the two as actually linked. The only one to express an explicit link said it was about, "Steering [the lesson] through, pacing it appropriately, [so] that they can meet this end point. Knowing when to move on and that whole thing about being directive".

5e is "making use of materials, resources and technical aids in such a way that they enhance learning". The most common interpretation of this criterion, explicitly mentioned by half those interviewed, was

that it refers first and foremost to use of the board, be it interactive or traditional. This is interesting because, as four participants observed, the board is not mentioned here at all, whereas it is mentioned in the descriptors for 2e. Six participants said they used 5e for effective and varied use of visuals and handouts, which five observed is linked to planning criterion 4c; four mentioned manipulation of listening equipment and three expressly mentioned not using technology just "for flash". Two commented that the reference to OHPs was rather outdated and could be removed. Only one tutor seemed to feel the criterion was generally not given enough weight and that creative candidates did not receive enough credit.

5f is "using appropriate means to make instructions for tasks and activities clear to learners". As mentioned above, there is noticeable overlap between this and 5b. Those who saw a clear separation between 5b and 5f tended to interpret 5b as choice of interaction patterns (or "pairwork" as two of them put it) and 5f as the wording of instructions. Although no specific techniques for checking understanding of instructions are mentioned here, nine tutors explicitly interpreted it as so-called "instruction-checking questions"; conversely, one tutor actually forbids this type of question and another was glad that Cambridge do not specify that they are to be used. The descriptor, "give an example or demonstration of the task if appropriate", which appears in both 5b and 5f, is more often seen as belonging in 5f. One explanation offered for the overlap between these two criteria was, "I think they [Cambridge] try to cover their behinds and so they kind of overdo it".

5g is "using a range of questions effectively for the purpose of elicitation and checking of understanding". This criterion is known in CELTA circles for being divisive. For some tutors (eleven in this study), 5g almost exclusively refers to so-called 'concept-checking questions' (or 'CCQs'), a method of checking understanding of language items seen by some to be essential and by others to be used with extreme caution, as just one of many methods (for more information about concept-checking questions see Dellar 2017; BBC/British Council; Florkowska 2018; Workman 2006). For tutors who favour 'CCQs', not using them will result in this criterion not being met, which is something others find shocking. Those less enamoured of the technique were most likely to raise objections to the wording of 5g, in four cases feeling that "using questions effectively" was something important in its own right, not necessarily or exclusively connected to eliciting and/or checking understanding. Three of these same tutors also pointed out that a lot of effective checking of understanding is done via coursebook exercises. Three tutors observed that "concept checking questions" are mentioned in 2e and a fourth wondered whether they fell under 2e or 5g depending at what point in the lesson they were used. It is also noticeable from some of the interviews that for a number of tutors the term 'CCQ' has been expanded to refer to a wider range of question types. There was some concern that *checking understanding of instructions* was also mentioned in the descriptors here. While most tutors seemed to ignore it and put it under 5f, two felt that if it was done using questions it came under 5g and if another method was used it was 5f. *Elicitation* also appears to be a term that is differently interpreted. While for many it simply means getting information from learners

rather than giving it, for others it implies a very specific technique, which again some felt was overused, to the extent that one tutor felt it would suffice to subsume it discreetly under 5c and not necessarily name it. The most unusual reaction to 5g, from a former Trinity tutor, was, "Why is it there? It's almost like it's been missed in the editing […] I'm not sure how necessary it is to have both [2e and 5g]".

To summarise, the main issue with 5g is that it is not clear whether it refers to use of questions generally, checking of understanding generally, or really to the more limited areas it appears to suggest. And if so, why is there no criterion for other uses of questions or for other methods of checking understanding?

5h is "providing learners with appropriate feedback on tasks and activities", where "feedback" seemed to mean four things to most people: allowing learners to compare answers in pairs; reporting on the content of discussions to the whole group; checking answers for factual accuracy; giving praise, comment and correction on language use, possibly including the shaping of emergent language. Within this, there were variations in priority and focus and sometimes one of the elements was subsumed elsewhere (such as pair checks under 5b, or language feedback under 2b). Some tutors felt feedback should always be given, while others felt good judgement could be knowing when to omit it. Five tutors mentioned the relationship between monitoring and feedback and two of those said it would make more sense for monitoring to come before feedback in the list of criteria, as it would in the classroom. Two tutors questioned the meaning of the word *appropriate* here and two felt some more concrete examples would be helpful. One thought, "There should be something about getting learners to listen to each other in feedback".

5i is "maintaining an appropriate learning pace in relation to materials, tasks and activities". Fourteen tutors said they felt "pace" was an issue when a lesson or activity dragged or when candidates rushed learners through things, with one adding that "there's inherent subjectivity in 'too long'"; only one explicitly said he rarely mentioned the criterion in a positive light, but this was perhaps implied by several others. Eight tutors explicitly linked pace to timing and three to the planning criterion for timing, 4h ("allocating appropriate timing for different stages in the lessons"). Five pointed out a link to monitoring (one descriptor is almost identical) and five to managing the learning process (5d). Two tutors objected to the inclusion of the descriptor "keep teacher talk to a minimum", a third said it was "controversial nowadays", and two others appeared puzzled by it. Three tutors felt it was there because of what one described as a "causal link" between excessive teacher talk and pace or timing issues. One tutor had a strong objection to the word *pace*, feeling that the point here was what he called *minds on*, which he considered very hard to evaluate.

5j ("monitoring learners appropriately in relation to the task or activity") showed a reasonably high level of consistency in interpretation, with nine tutors pointing out that they expected monitoring to be done in more than a superficial way and for candidates to use what they noticed to help with decisions

related to 2b, 2d, 5d, 5i and/or 5h. Two tutors felt that monitoring and feedback should be explicitly linked in the criteria, with wording such as, "Providing learners with appropriate feedback based on monitoring". Response to learners and a link to rapport were also mentioned by several here. The only controversial issue appeared to be whether monitoring from a distance was acceptable or whether candidates were always expected to get physically close to students.

5k ("beginning and finishing lessons on time and, if necessary, making any relevant regulations pertaining to the teaching institution clear to learners") was not considered important by many tutors and seven said they rarely if ever referred to it, one saying it was "to standard by default". Nine said they took responsibility themselves for most of it, although one said it was an important part of professionalism and should be down to the candidates, while another said it would be "good for them" if it were. Two said that it was linked to pace, in the words of one, "5i is instructive and 5k is punitive". One felt that rewording along the lines of "finishing lessons on an appropriate note" would be more useful, to help prevent abrupt lesson endings.

5   Discussion and conclusions

Aristotle said, "It is the mark of an educated man to require, in each kind of inquiry, just so much exactness as the subject admits of" (1893: I. 3, 4), and it is this which makes both assessing teaching and analysing and comparing people's interpretations of how they themselves assess teaching a challenge (see also UCLES 2018: 15). Attempting to define *consistency* of interpretation made it clear just how difficult this is and, perhaps, why Thaine did not even try. The findings of this study appear to suggest strong inconsistency of interpretation of the CELTA assessment criteria for teaching, but this is to a certain extent also an interpretation on the part of the researcher, who has been very cautious about assuming that any two people are, in fact, saying the same thing. It has also been a personal decision to judge when differences in 'operationalisation' are significant enough to be considered different 'interpretations'. Consequently, only two criteria have been deemed to be interpreted with a truly high level of consistency. Even here, however, this does not mean that all is well: criterion 2f revealed a high level of consistency in being considered pointless! The same thing is to be seen among the criteria which showed a 'reasonable' level of consistency: of the five, two (5a and 5k) are generally ignored or given low priority. As seen in 3.1 above, Thaine found that inconsistency usually stemmed from overlap between criteria or over-generalisation; in the current study, some additional issues can be identified, and the question of whether overlap is either avoidable or undesirable should also be addressed. It emerged as well that, while tutors generally appear satisfied that the criteria are useful, there is a strong feeling that they need reviewing, in some cases rewording, that some areas lack sufficient exemplification, and that some important elements of teaching are missing or at least not explicitly mentioned. Let us look first at the causes of inconsistency:

- Overlap is still present. Following Thaine's study, descriptors were added with the intention of making the criteria themselves more transparent and of exemplifying them. It seems, however, that in some cases the opposite result has been achieved and the descriptors actually confuse or obscure rather than elucidate. This is the case for criteria 1a – 1c, where two descriptors are identical for each, one of which explicitly refers to planning rather than teaching and one of which overlaps with criterion 5b. There are also almost-identical descriptors in 5f and 5g, as well as 5i and 5j. More general overlap is seen in the wording of descriptors for 2a, 5f and 5i; 5d, 5i and 5j. This is undoubtedly deliberate but it is not explained. This is explored further in 5.1 below.

- Too many or too dissimilar elements under one criterion heading. This was most frequently perceived in the case of 2e but was also commented on with regard to Topic 3 ("Input is so important and they're lumped together, reading and listening"). Interestingly, while some people felt strongly enough about 2e to 'rebel' and divide the criterion up for themselves somehow, this does not seem to occur with Topic 3, where the conflation is most often seen as tacit permission to ignore writing skills and/or teach reading and listening in the same way.
- Lack of examples, imbalance in or vagueness of descriptors. The level of detail in the descriptors is noticeably inconsistent. 2e, for instance, specifies concrete techniques which could be used ("e.g. finger highlighting, highlighting on the board"), and 5e specifies (rather quaintly), "OHP, video, projector or sound system". Other criteria, however, opaquely reference "procedures you have learnt on the course". This is particularly odd in the case of 5c, which is all about techniques and lesson types and yet not one concrete technique or lesson type is mentioned (Cambridge English 2018b: 28 - 29). While one tutor suggested this means, "We are free to interpret that in our own way", it is unclear whether this is really the case and there are some definite conventions; for instance, the same tutor suggested focusing on connected speech as something to do in preparation for a listening task, and his colleague replied, "I'm sure some trainers would say – well, they have said to me, you know, 'What do you mean focusing on connected speech before a listening task, they can't do that on the CELTA'".

Other areas which caused concern and in some cases also contributed to differences in interpretation or operationalisation were:

- Illogical ordering. The criteria are organised into broad topic areas but within those areas are only partially in a logical order. This perturbed tutors most in Topic 5, which starts logically enough with 5a being about arranging the classroom and progresses to 5k about finishing the lesson on time. However, between these, several tutors found it illogical that 5b and 5f covered so much similar ground and yet were so far apart, and others felt that it made no sense to have monitoring (5j) after feedback (5h), when feedback is supposed to be given in response to and based on the results of monitoring.
- Lack of weighting. In the words of one tutor, "All criteria are equal but some are more equal than others". There is, in fact, no explicit criteria weighting and yet, for most tutors, an enormous amount of emphasis is placed on 2e and 5g, while criteria like 1a-1c, 2f, 5a and 5k are often considered met by default. This and the previous point were noticed by Galaczi and Swabey and one of their conclusions is, "A way forward would be to simplify the set of assessment criteria through clear organisation and weighting" (2015: 122).
- No reference to the stage of the course. The *CELTA 5* booklet refers several times to the standard "expected at this stage of the course" (e.g. Cambridge English 2018b: 14) and many

centres include this wording in their feedback *pro forma*. Cambridge give no clear indication of what they mean by this, which has led to two very different interpretations. One, which is often formalised by individual centres and is sometimes called 'the pyramid', is where candidates are expected to meet a limited number of specified criteria in the first week of the course, which is then either built upon or, in some cases, even moved on from each week. This is so set-in-stone in some centres that newer tutors who find it illogical sometimes question it without realising it is not a Cambridge requirement[3]: "What I think really matters is contrasting the first contact […with], halfway through the course, meeting a new group of learners and still being able to establish this rapport […] And [yet] this rapport criterion [1d] disappears because of the fact that we are elsewhere on the journey, right?". The second interpretation is that candidates should meet all the criteria, as and when they are relevant to their lessons, throughout the course but should be meeting them more convincingly and consistently in the fourth week than the first. One tutor interviewed felt that the very concept of "for the stage of the course" made no sense.

- Controversial wording. Several criteria provoked strong reactions from some tutors on account of the wording, most often of the descriptors but occasionally of the criteria themselves. One key reason was because they felt it led to over-emphasis on or desirability of certain behaviours, was too reductive in key areas ("I don't have a problem with teacher talk, as long as it's *quality* teacher talk"), or suggested false dichotomies (e.g. 'learner-centred' versus 'teacher-fronted'). In the case of the descriptors, tutors generally chose to ignore the ones they did not like ("To me, 5g is only checking understanding of language, [checking understanding of instructions] come[s] under 5f"). In the case of 5g, the juxtaposition of *questions, elicitation* and *checking of understanding* irked tutors who felt that the three did not always go together and that *use of a range of questions* and *checking understanding* could usefully be two separate criteria.

- Areas missing. Several tutors wanted more explicit reference to effective use of the voice, response to learner needs (including deviating from the plan as a result, although one tutor said this was not to be expected at CELTA level), dealing with emergent language, fostering learner autonomy, authenticity of discourse with learners. Several also felt that, although Topic 5 includes the word *professionalism* in the heading, true elements of professionalism within the criteria were insufficient.

5.1   The significance of overlap between criteria

---

[3] It could perhaps be seen as implicit in the information given on p. 8 of the *CELTA 5* booklet: "In the early stages of the course, you are assessed against *specific assessment criteria* and partial performance related to what can be expected at that stage of the course" (italics added), although there is no indication of which criteria these should be. See also Pulverness 2015: 24.

Thaine clearly identified overlap between criteria as a problem (2004: 342 – 343). However, several of the current criteria share identical or almost-identical descriptors and cover some of the same ground. It seems extremely unlikely that this is accidental, yet no explanation is given. Some tutors interviewed found the overlap confusing or frustrating, while others had their own explanations for it: it was to show how things are linked; it was to emphasise just how important some areas are; it was to make sure key points are definitely covered somewhere. With a few exceptions, newer tutors fell into the former category, asking questions to the effect of, "Which criterion am I supposed to use for this?", while more experienced tutors simply decided for themselves, although these decisions were not always the same. The impression given was, in the words of one, "[I don't think overlap/repetition] is a problem, as long as you explain to the candidates which one you're using and why" – although the same tutor admitted that this has led to some disagreement with colleagues. It would appear that, despite some suggestion in the literature that overlap is problematic (Bingham 2002: 11; Thaine 2004), Cambridge may have decided that, in the case of the CELTA, it is unavoidable or even advantageous, perhaps for the reasons some of the tutors interviewed surmised.

5.2   Advantages, disadvantages and limitations of the present situation

The main advantage of the present set of criteria, the way they are presented and the way tutors are trained (internally, by each other), is clearly that there is some scope for individuality, variety and, importantly, adaptation to context. One key disadvantage is that the way the criteria are currently worded and used by the majority of tutors has perpetuated the 'myth' cited in 2.2.1 above that there is a standard lesson expected on the CELTA, which risks undermining the aforementioned advantages. In fact, there could be said to be two such 'standard lessons', one for grammar, which follows the PPP framework seen above (2.2.1; 4.4.2) and one for reading/listening, which follows a popular 'top-down' framework, which has been convincingly challenged as inadequate both for reading (Swan and Walter 2017; Thornbury 27.11.2011 and n.d.; Nuttall 2005: 30) and listening (Field 2008: 27 – 33; Rost 1990: 9 – 10; Swan and Walter 2017). With regard to the PPP grammar lesson framework, it is of some concern that tutors vary so much on how they expect it to be used: the difference most frequently mentioned was whether they accepted 'omission' of the so-called 'freer practice' (Production) stage but, perhaps more relevant, is that some tutors insist on one individual candidate completing all stages of a PPP lesson in one 40- or 45-minute slot, while others divide the Ps, as it were, between the candidates, with the first teaching perhaps a reading lesson which establishes a context for the language, the second dealing with the Presentation and ('controlled' or 'restricted') Practice stage and the third having a full 40 minutes for substantial 'freer practice'. Clearly a teacher who has to pack all of these things into a third of the time is not going to do so as thoroughly as one who has only one 'P' to cover, and is more likely to produce a display lesson rather than really teach the learners in front of them. At the same time, planning the whole process is perhaps more likely to engrain the 'recipe' in their mind and help them see the connection between the parts. A teacher with just one 'P' to cover may do so more thoroughly but is dependent on the teacher before them or under pressure from the

teacher after them to treat their 'P' in a satisfactory manner, and may fail to clearly see how the three parts of the lesson are designed to form a coherent whole. There is also a lot of variation in the extent to which 'alternatives' to the PPP model are expected, or even accepted, by tutors. A further disadvantage of the current situation, mentioned by several tutors, is that the summative assessment side of the CELTA dominates and becomes an exercise in what was variously described as "box-ticking", "bean-counting" or "jumping through hoops". This washback effect, which is mentioned in the literature (Malderez and Wedell 2007: 158; Norton 2004: 689; Ochieng Ong'ondo and Borg 2011: 523), is perhaps inevitable, given that candidates have 120 hours in which to show that they can meet 41 criteria satisfactorily, but it is clearly something that some tutors dislike and try to minimise. One tutor said, "I think, deep down probably the CELTA's very behaviourist in its approach and a lot of trainees, all they can cope with is, 'What do I have to do to get through? [...] I just want to display these things so that you can tick the boxes'. And it's very difficult to go beyond that" (see also Gebhard 1990: 158). Thus, many tutors do expect candidates to learn lesson 'recipes' and follow them, regardless of whether they are suitable for the learners participating in the Teaching Practice lessons (tellingly referred to by some as 'guinea pigs'), which results in display lessons rather than real teaching, something which seems to contradict the desire expressed by so many tutors (and criterion 1a) that candidates show awareness of learner needs. While the nature of this study shows that this is clearly a concern of some tutors, a study by Brandt published in 2006 suggests that some candidates also feel this way (: 356).

5.3   The way forward

Despite pages and pages of course documentation, a substantial syllabus and several handbooks, as well as an annual standardisation process, the question still coming from many tutors is, "Is this what Cambridge want?". It is certainly positive that some flexibility is contained within the assessment criteria but there is clearly a general feeling that a little more clarity regarding both the meaning and desired use of the criteria would be appreciated. The results of this study, and indeed a number of the participants themselves, suggest that the CELTA criteria are in need of a more substantial overhaul than the minor changes effectuated in 2017. That Cambridge are aware of this is also apparent from Galaczi and Swabey's study published in 2015. Both their study and this one indicate that a more logical ordering of the criteria as well as some sort of weighting would be useful. In addition, this study suggests a need for some rewording, clearer exemplification via the descriptors, and in some cases restructuring to reduce excessive overlap and ambiguity, and to ensure that individual criteria do not attempt to cover too much or too little. This is obviously a major undertaking and it is beyond the scope of this study to suggest how it might be done. However, it is pertinent to consider whether revising the criteria is enough in itself. The literature shows that not only should assessment criteria be regularly reviewed but that assessors (in this case CELTA tutors as well as those in the role entitled Assessor) need to be suitably trained to use them. It could be questioned whether this is currently the case on the CELTA. As seen above (3.2.1; 5.2), tutors are trained internally by each other, with a

portfolio of work submitted to Cambridge and the Assessor's visit to a course which the new tutor is shadowing being the only external elements. In order to train others, a tutor simply needs approval from Cambridge which, when granted, is then valid for life. There is no requirement for continued professional development, simply an annual online standardisation procedure, which involves all tutors around the world watching a recorded lesson and marking some written assignments. They submit their overall grade for approval from Cambridge and if too few of their grades coincide with the official grades they are required to do more standardisation tasks. While they are encouraged to take notes on the lessons and essays and are later given the official feedback comments to read, none of this part is checked in any way; only the overall grading counts (Galaczi and Swabey 2015: 131 – 135). It does seem that this is Cambridge's main concern and that 'standardisation' means agreeing on overall grades. Whether within that scope candidates are really receiving comparable training and whether the ideas underpinning this training are moving with the times is questionable and probably worth investigating if the qualification is to maintain its value into the future (ELTConcourse n.d.: 2 - 6).

(WORD COUNT 15, 971)

Alderson, J., Clapham, C. and Wall, D. (1995). *Language test construction and evaluation.* CUP

Anderson, J. (2017). "A potted history of PPP with the help of *ELT Journal*" in *ELT Journal* 71/2. OUP

Anderson, J. (2018). "'Buying in' to communicative language teaching: the impact of 'initial' certification courses on the classroom practices of experienced teachers of English" in *Innovation in language learning and teaching,* 1 -14. Routledge.

Aristotle, The Nichomachean Ethics of Aristotle, trans. F.H. Peters, M.A. 5th edition. London: Kegan Paul, Trench, Truebner & Co., 1893. Available at https://oll.libertyfund.org/titles/903#Aristotle_0328_19. Accessed 13.4.2019.

Bailey, K. (1990). "The use of diary studies in teacher education programs" in *Second language teacher education.* Richards, J. and Nunan, D. (eds.) CUP

Bailey, K., Bergthold, B., Braunstein, B., Jagodzinski Fleischman, N., Holbrook, M., Tuman, J., Waissbluth, X. and Zambo, L. (1996) "The language learner's autobiography: examining the 'apprenticeship of observation'" in *Teacher learning in language teaching.* Freeman, D. and Richards, J. (eds.). CUP

BBC / British Council teaching English website available at https://www.teachingenglish.org.uk/article/checking-understanding accessed 25.8.2018

Bell English (2017) available at: https://www.bellenglish.com/sites/default/files/public/uploads/Brochures/Bell%20IATEFL%20Leaflet_2017.pdf accessed 6.1.2019

Bingham, R. (2002). *Learning outcomes and assessment criteria.* Sheffield Hallam University

Black, P. and Wiliam, D. (1998). "Assessment and classroom learning" in *Assessment in Education: principles, policy & practice* 5/1, 7 – 74; doi: 10.1080/0969595980050102. Routledge

Borg, S. (2009). "Language teacher cognition" in *The Cambridge guide to second language teacher education*. Burns, A. and Richards, J. (eds.). CUP

Brandt, C. (2006). "Allowing for practice: a critical issue in TESOL teacher preparation" in *ELT Journal* 60/4 October 2006, 355 – 364; doi: 10.1093/elt/ccl026. OUP

Brandt, C. (2008). "Integrating feedback and reflection in teacher preparation" in *ELT Journal* 62/1 January 2008, 37 – 46; doi: 10.1093/elt/ccm076. OUP

Burns, A. and Richards, J. (2009). *The Cambridge guide to second language teacher education.* CUP

Burton, J. (2009). "Reflective practice" in *the Cambridge guide to second language teacher education.* Burns, A. and Richards, J. (eds.). CUP

Byrne, D. (1986). *Teaching oral English (new edition).* Harlow: Longman

Cambridge English (2015). *CELTA overview.* UCLES. Available at
https://www.cambridgeenglish.org/Images/272250-celta-overview.pdf

Cambridge English (2016) *Trainer Framework.* UCLES. Available at
http://www.cambridgeenglish.org/teaching-english/cambridge-english-trainer-framework/
accessed 31.12.18

Cambridge English (2014) *Teaching Framework.* Available at
https://www.cambridgeenglish.org/teaching-english/cambridge-english-teaching-framework/
accessed 1.1.19

Cambridge English (2017). *CELTA administration handbook.* UCLES.

Cambridge English (2018a) *CELTA syllabus and assessment guidlines* (5th ed.) UCLES. Available at
www.cambridge.org/celta accessed 11.7.18

Cambridge English (2018b). *CELTA candidate record booklet, CELTA 5.*

Crooks, T. (2004). *Tensions between assessment for learning and assessment for qualifications.* Paper
presented at the Third Conference of the Association of Commonwealth Examinations and
Accreditation Bodies (ACEAB), Nadi, Fiji, 8-12 March 2004. Accessed online 21.6.2018

Danielson, C. (2012). "Observing classroom" in *Educational leadership* Nov. 2012

Danielson, C. (2014). *Charlotte Danielson's framework for teaching.* Available at
http://www.k12.wa.us/TPEP/Frameworks/Danielson/2011_Danielson_SMART_Card.pdf -
accessed 2.1.2019

Darling-Hammond, L. and Snyder, J. (2000). "Authentic assessment of teaching in context" in
Teaching and teacher education 16, 523 – 545

Delaney, J. (2015). "The 'dirty mirror' of reflective practice: assessing self- and peer-assessment on a
CELTA course" in *Studies in language testing 42: assessing language teachers' professional
skills and knowledge.* Wilson, R. and Poulter, M. (eds.) CUP

Dellar, H. (2017). "On the over-use of concept-checking questions: part 1" at
https://www.lexicallab.com/2017/10/on-the-over-use-of-concept-checking-questions accessed
27.8.18

Douglas, J. (28.12.2018). personal correspondence

EAQUALS European Profiling Grid (2013) available at https://www.eaquals.org/wp-content/uploads/The_EPG_-_PDF_publication_final.pdf accessed 1.1.2019

Ellis, R. (2009). "SLA and Teacher Education" in *The Cambridge guide to second language teacher education.* Burns, A. and Richards, J. (eds.). CUP

Ellis, R. and Shintani, N. (2014) *Exploring language pedagogy through second language acquisition research.* London and New York: Routledge

ELTConcourse (2018) at
http://www.eltconcourse.com/training/initial/celta/syllabuscelta/unpacking_celta_syllabus.html, accessed 10.3.2019

ELTConcourse (n.d.) "Evolution and decay in ELT methodology and training at
http://www.eltconcourse.com/articles/articles1/ELT_Concourse_Evolution_and_decay_in_ELT_methodology_and_training.pdf accessed 13.7.2019

Ferguson, G. and Donno, D. (2003). "One-month teacher training courses: time for a change?" in *ELT Journal* 57/1 January 2003, 26 – 33. OUP

Field, J. (2008). *Listening in the language classroom.* CUP

Florkowska, K. (2018). *The use of Concept Checking Questions in the EFL classroom: a reflection on the CELTA experience.* Available at
https://englishagenda.britishcouncil.org/sites/default/files/attachments/karolina_florkowska_kingston_university_london_dissertation.pdf. Accessed 25.8.2018

Freeman, D. (1990). "Intervening in practice teaching" in *Second language teacher education.* Richards, J. and Nunan, D. (eds.). CUP

Freeman, D. (1996). "The 'unstudied problem': research on teaching learning in language teaching" in Freeman, D. and Richards, J. (eds.). *Teacher learning in language teaching.* CUP

Freeman, D. and Richards, J. (eds.) (1996). *Teacher learning in language teaching.* CUP

Freeman, D., McBee Orzulak, M., Morrisey, G. (2009). "Assessment in second language teacher education" in *The Cambridge guide to second language teacher education.* Burns, A. and Richards, J. (eds.). CUP

Gaies, S. and Bowers, R. (1990). "Clinical supervision of language teaching: the supervisor as trainer and educator" in *Second language teacher education.* Richards, R. and Nunan, D. (eds.). CUP

Galaczi, E. and Swabey, M. (2015). "Capturing the ephemeral: standardising the assessment of teaching practice" in *Studies in language testing 42: assessing language teachers' professional skills and knowledge.* Wilson, R. and Poulter, M. (eds.) CUP

Gebhard, J. (1990). "Models of supervision: choices" in *Second language teacher education.* Richards, J. and Nunan, D. (eds.). CUP

Gebhard, J. (2009). "The practicum" in *The Cambridge guide to second language teacher education.* Burns, A. and Richards, J. (eds.). CUP

Graves, K. (2009). "The curriculum of second language teacher education" in *The Cambridge guide to second language teacher education.* Burns, A. and Richards, J. (eds.). CUP

Hawkey, R. and Milanovic, M. (2013). *Cambridge English exams – the first hundred years: a history of English language assessment from the Univerisity of Cambridge, 1913 – 2013.* CUP

Hobbs, V. (2013). "'A basic starter pack': the TESOL certification as a course in survival" in *ELT Journal* 67/2 April 2013, 163 - 174; doi: 10.1093/elt/ccs078. OUP

Holliday, A. (1998). "Evaluating the discourse: the role of applied linguistics in the management of evaluation and innovation" in *Managing evaluation and innovation in language teaching: building bridges.* Rea-Dickins, P. and Germaine, K. (eds.). Harlow: Longman

Howatt, A. with Widdowson, H. (2004). *A history of English language teaching. 2nd ed.* OUP

Johnson, K. (2009). "Trends in second language teacher education" in *The Cambridge guide to second language teacher education.* Burns, A. and Richards, J. (eds.). CUP

Johnson, K. and Golombek, P. (2018). "Informing and transforming language teacher education pedagogy" in *Language teaching research*, 1-12; doi: 10.1177/1362168818777539. SAGE

Katz, A. and Snow, M. (2009). "Standards and second language teacher education" in *The Cambridge guide to second language teacher education.* Burns, A. and Richards, J. (eds.). CUP

Korthagen, F. (2017). "Inconvenient truths about teacher learning: towards professional development 3.0" in *Teachers and training* 23/4, 387 – 405; doi: 10.1080/13540602.2016.1211523

Lange, D. (1990). "A blueprint for a teacher development program" in *Second language teacher education.* Richards, J. and Nunan, D. (eds.). CUP

Malderez, A. and Wedell, M. (2007). *Teaching teachers: processes and practices.* London: Continuum

Morgan, M. (2015). "Developing and assessing English language teacher trainers" in *Studies in language testing 42: assessing language teachers' professional skills and knowledge.* Wilson, R. and Poulter, M. (eds.) CUP

Norton, L. (2004). "Using assessment criteria as learning criteria: a case study in psychology" in *Assessment & evaluation in higher education* 29/6, 687 -702. Doi: 10.1080/0260293042000227236. Routledge

Nunan, D. (1990). "Action research in the language classroom" in *Second language teacher education.* Richards, J. and Nunan, D. (eds.). CUP

Nuttall, C. (2005). *Teaching reading skills in a foreign language* (2005). Oxford: MacMillan

Ochieng' Ong'ondo, C. and Borg, S. (2011). "'We teach them plastic lessons to please them': the influence of supervision on the practice of English language student teachers in Kenya" in *Language teaching research*, 15/4, 509 – 528; doi: 10.1177/1362168811412881. Sage

Poulter, M. (2007). "Cambridge ESOL teacher training and development – future directions" in *Cambridge ESOL research notes* 29 / August 2007, 2 - 4. Accessed online 21.6.2018

Pulverness, A. (2015). "A brief history of Cambridge English Language Assessment teaching qualifications" in *Studies in language testing 42: assessing language teachers' professional skills and knowledge.* Wilson, R. and Poulter, M. (eds.) CUP

Randall, M. and Thornton, B. (2001). *Advising and supporting teachers.* CUP

Rea-Dickins, P. and Germaine, K. (1992). *Evaluation.* OUP

Richards, J. (1990). "The dilemma of teacher education in second language teaching" in *Second language teacher education.* Richards, J. and Nunan, D. (eds.). CUP

Richards, J., Ho, B. and Giblin, K. (1996). "Learning how to teach in the RSA Cert" in *Teacher learning in language teaching.* Freeman, D. and Richards, J. (eds.). CUP

Richards, J. (1998). *Beyond training.* CUP

Richards, J. and Ho, B. (1998). "Reflective thinking through journal writing" in *Beyond training.* CUP

Richards, J. (2002). "Theories of teaching in language teaching" in *Methodology in language teaching: an anthology of current practice.* Richards, J. and Renandya, W. CUP

Richards, J. and Farrell, T. (2005). *Professional development for language teachers: strategies for teacher learning.* CUP

Roberts, J. (1998). *Language teacher education.* London: Arnold

Rossner, R. (2009). "Developing common criteria for comparison and assessment in language teacher education" in *Cambridge ESOL research notes,* issue 38, November 2009. UCLES

Rost, M. (1990). *Listening in language learning.* London: Longman

Shinkfield, A. and Stufflebeam, D. (1995). *Teacher evaluation: Guide to effective practice.* Boston/Dordrecht/London: Kluwer

Schön, D. (1987). *Educating the reflective practitioner.* San Francisco: Jossey-Bass

Shulman, L. (1986). "Those who understand: knowledge growth in teaching". In *Educational researcher,* 15/2, 4 -14. JSTOR.

Smith, D. (1996). "Teacher decision making in the adult ESL classroom" in *Teacher learning in language teaching*. Freeman, D. and Richards, J. (eds.). CUP

Strong, M., Gargani, J. and Hacifazlioglu, Ö. (2011). "Do we know a successful teacher when we see one? Experiments in the identification of effective teachers" in *Journal of teacher education* 62/4, 367 – 382; doi: 10.1177/0022487110390221. SAGE

Sulaimani, A. and Elyas, T. (2015). "A critical review of the CELTA syllabus within the context of Saudi Arabia". De Gruyter open; doi: 10.1515/rjes-2015-0003

Swan, M. and Walter, C. (2017) "Misunderstanding comprehension" in *ELT Journal* 71/2 April 2017, 228 – 236. OUP

Thaine, C. (2004). "The assessment of second language teaching" in *ELT Journal* 58/4 October 2004, 336 – 345. OUP

Thaine, C. (14.2.2019). Personal correspondence

Thornbury, S. (n.d.) at http://www.onestopenglish.com/methodology/teaching-articles/debates/the-end-of-reading/debate-the-end-of-reading-part-three-some-practical-suggestions/144671.article. Accessed 19.4.2019

Thornbury, S. (27.11.2011) at https://scottthornbury.wordpress.com/2011/11/27/g-is-for-gist/ accessed 27.4.2019

Thornbury, S. and Watkins, P. (2007). *The CELTA course. Trainer's manual.* CUP

Trinity College London (2019). https://www.trinitycollege.com/site/?id=201. Accessed 20.6.19

UCLES (2018). https://www.cambridgeassessment.org.uk/Images/cambridge-approach-to-assessment.pdf. Accessed 18.7.19

UCLES (2019a). https://www.cambridgeenglish.org/find-a-centre/find-a-teaching-centre/. Accessed 20.6.19

UCLES (2019b). https://www.cambridgeenglish.org/why-choose-us/producing-exams/. Accessed 18.7.19

Wallace, M. (1991). *Training foreign language teachers: a reflective approach.* CUP

Watkins, P., Harris, B. and Pulverness, A. (2015). "Using international teacher education programmes in local contexts" in *Studies in language testing 42: Assessing language teachers' professional skills and knowledge.* Wilson, R. and Poulter, M. (eds.). CUP

Wilson, R. and Poulter, M., eds. (2015). *Studies in language testing 42: assessing language teachers' professional skills and knowledge.* CUP.

Workman, G. (2006) *Concept Questions And Time Lines*. Chadburn Publishing.

Wright, T. (2009). "'Trainer development': professional development for language teacher education" in *The Cambridge guide to second language teacher education.* Burns, A. and Richards, J. (eds.) CUP

Wright, T. and Bolitho, R. (2007). *Trainer development.* Lulu.com

Appendix A

<u>Topic 1 – learners and teachers and the teaching and learning context</u>

1a   teaching a class with an awareness of the needs and interests of the learner group

- find out from learners and peers about the needs and interests of learners
- use this information for selecting materials and activity types where appropriate
- use this information when setting up pair and group work and dealing with learners in open class where appropriate

1b   teaching a class with an awareness of learning preferences and cultural factors that may affect learning

- find out from learners and peers about the cultural backgrounds oflearners
- use this information for selecting materials and activity types where appropriate
- use this information when setting up pair and group work and dealing with learners in open class where appropriate

1c   acknowledging, when necessary, learners' backgrounds and previous learning experiences

- find out from learners and peers about the learning backgrounds of learners
- use this information for selecting materials and activity types where appropriate
- use this information when setting up pair and group work and dealing with learners in open class where appropriate

1d   establishing good rapport with learners and ensuring they are fully involved in learning activities

- build a positive classroom atmosphere
- interact naturally with learners before, during and after the lesson
- maintain eye contact
- ensure that learners are involved in the lesson during teacher-fronted and learner-centred stages of the lesson.

<u>Topic 2 - language analysis and awareness</u>

2a   adjusting their own use of language in the classroom according to the learner group and the context

- use simple language to give instructions and when explaining
- keep your simplified language natural
- allow learners opportunity to speak by keeping teacher talk to an appropriate level

2b   identifying errors and sensitively correcting learners' oral and written language

- show an awareness of student errors
- correct learners' language sensitively during controlled oral practice activities
- give feedback on oral errors after a communicative activity
- correct learners' language sensitively during controlled written practice activities
- correct freer written tasks set in class or set for homework

2c   providing clear contexts and a communicative focus for language

- provide a context for language by means of text, situation or task using visual aids and realia as appropriate
- ensure there is a clear link between the context and the target language
- ensure that the context provides learners with sufficient opportunity for communicative practice

2d   providing accurate and appropriate models of oral and written language in the classroom

- choose natural examples of language from context
- ensure new language models are natural and accurate when drilling
- highlight the target language clearly
- ensure language used on the board and on worksheets is correct in terms of spelling and punctuation

2e   focusing on language items in the classroom by clarifying relevant aspects of meaning and form (including phonology) to an appropriate degree of depth

- clarify the meaning of language in language-based lessons by using one of the ways you have learnt on the course, e.g. concept questions, timelines or a learner-centred task
- clarify the form of language in language-based lessons by using one of the ways you have learnt on the course, e.g. using the board or a learner-centred task
- clarify the pronunciation of language in language-based lessons in one of the ways you have learnt on the course, e.g. finger highlighting, highlighting on the board

2f   showing awareness of differences in register

- show awareness of formal, neutral and informal language
- show awareness of how language changes according to the different contexts in which it is used

2g   providing appropriate practice of language items

- provide as much practice in context as possible
- ensure the practice is appropriate to the target language
- stage practice activities logically

Topic 3 – language skills: reading, listening, speaking and writing

3a   helping learners to understand reading and listening texts

- follow teaching procedures you have learnt on the course for a receptive skills-based lesson
- ensure an appropriate focus on developing receptive skills and subskills

3b   helping learners to produce oral and written language

- follow teaching procedures you have learnt on the course for a speaking skills-based lesson
- ensure an appropriate focus on developing speaking skills and subskills
- ensure a communicative focus in speaking activities
- provide learners with opportunities to practise writing in language-focused and skills lessons
- ensure an appropriate focus on practising writing skills and subskills

Topic 5 – developing teaching skills and professionalism

5a   arranging the physical features oft he classroom appropriately for teaching and learning, bearing in mind safety regulations of the institution

- arrange the furniture and equipment in the classrom to suit different types of activity

5b   setting up and managing whole class and/or group or individual activities as appropriate

- give clear instructions for pair, group, individual and plenary work
- organise the learners in pair, group, individual and plenary work
- give an example or demonstration of the task if appropriate

5c   selecting appropriate teaching techniques in relation to the content of the lesson

- recognise different lesson types (skills based, language focus based) and follow teaching procedures you have learnt on the course to achieve the aims of different types of lesson

5d managing the learning process in such a way that the lesson aims are achieved

- ensure that the activities and tasks help achieve the aim of the lesson
- ensure there is an appropriate balance between teacher-fronted and learner-centred activities
- be sufficiently directive when appropriate
- keep a low profile when appropriate
- know when to intervene or not

5e making use of materials, resources and technical aids in such a way that they enhance learning

- use games, puzzles, pictures, realia, audio material to help learners learn and to provide practice
- use technical aids (OHP, video, projector or sound system) so that they are clear to all

5f using appropriate means to make instructions for tasks and activities clear to learners

- use simple language to give instructions for tasks and activities
- give instructions at an appropriate stage of the lesson
- give an example or demonstration of the task if appropriate
- check that learners have understood instructions for tasks and activities

5g using a range of questions effectively for the purpose of elicitation and checking of understanding

- use questions for
  - setting context
  - building up information
  - assessing learners' prior knowledge
  - checking meaning of language items
  - checking understanding of instructions

5h providing learners with appropriate feedback on tasks and activities

- give learners time to check the answers to tasks in pairs
- provide feedback on both the content of activities and the language used in them
- use a variety of techniques in order to give feedback on activities

5i maintaining an appropriate learning pace in relation to materials, tasks and activities

- keep teacher language and explanation to a minimum
- allow time for learners to complete tasks without allowing activities to go on too long
- be aware of when learners are ready to move on to the next stage of the lesson

5j monitoring learners appropriately in relation to the task or activity

- listen to learners attentively but unobtrusively during stages of the lesson
- know when to intervene in learner-centred activities
- ensure that your attention is spread evenly amongst the learners
- know when to move on to the next stage of the lesson

5k beginning and finishing lessons on time and, if necessary, making any relevatn regulations pertaining to the teaching institution clear to learners

- ensure that you are in the classroom in good time to begin your lesson on time
- ensure that your materials are prepared in good time to begin your lesson on time
- ensure that you finish your lesson on time and that you do not exceed your allotted time
- ensure that learners are aware of start and finish time as required
- ensure that you pass on any relevant administrative information to learners when required

Appendix B

Levels of consistency of interpretation of the criteria found in this study. The criteria appear here without the descriptors; these can be found in Appendix A.

** = high consistency (over 25 participants had the same interpretation)

* = reasonable consistency (20 – 24 participants had the same interpretation)

= low consistency (fewer than 20 participants had the same interpretation OR there were at least two very different interpretations)

TOPIC 1 – LEARNERS AND TEACHERS AND THE TEACHING AND LEARNING CONTEXT

#1a   teaching a class with an awareness of the needs and interests of the learner group

#1b   teaching a class with an awareness of learning preferences and cultural factors that might affect learning

#1c   acknowledging, when necessary, learners' backgrounds and previous learning experiences

*1d   establishing good rapport with learners and ensuring they are fully involved in learning activities

TOPIC 2 – LANGUAGE ANALYSIS AND AWARENESS

**2a   adjusting their own use of language in the classroom according to the learner group and the context

#2b   identifying errors and sensitively correcting learners' written and oral language

#2c   providing clear contexts and a communicative focus for language

#2d   providing accurate and appropriate models of oral and written language in the classroom

#2e   focusing on language items in the classroom by clarifying relevant aspects of meaning and form (including phonology) to an appropriate degree of depth

**2f   showing awareness of differences in register

#2g   providing appropriate practice of language items

TOPIC 3 – LANGUAGE SKILLS: READING, LISTENING, SPEAKING AND WRITING

#3a   helping learners to understand reading and listening texts

#3b   helping learners to produce oral and written language

TOPIC 5 – DEVELOPING TEACHING SKILLS AND PROFESSIONALISM

*5a   arranging the physical features of the classroom appropriately for teaching and learning, bearing in mind safety regulations of the institution

#5b   setting up and managing whole class and/or group or individual activities as appropriate

#5c   selecting appropriate teaching techniques in relation to the content of the lesson

#5d   managing the learning process in such a way that lesson aims are achieved

*5e   making use of materials, resources and technical aids in such a way that they enhance learning

#5f   using appropriate means to make instructions for tasks and activities clear to learners

#5g   using a range of questions effectively for the purpose of elicitation and checking of understanding

#5h   providing learners with appropriate feedback on tasks and activities

#5i   maintaining an appropriate learning pace in relation to materials, tasks and activities

*5j   monitoring learners appropriately in relation to the task or activity

*5k   beginning and finishing lessons on time and, if necessary, making any relevant regulations pertaining to the teaching institution clear to learners