



Cambridge Assessment  
English

Research Notes

# Context and quality assurance in language testing

Issue 80  
July 2021

ISSN: 1756-509X



**Cambridge Assessment  
English**

## **Research Notes**

**Issue 80/July 2021**

### **Senior Editor and Editor**

Siân Morgan, Senior Research Manager, Research and Thought Leadership Group,  
Cambridge Assessment English

John Savage, Publications Assistant, Research and Thought Leadership Group,  
Cambridge Assessment English

Typeset in the United Kingdom by Abound Creative

# Research Notes: 80

## CONTENTS

---

**2** Editorial

---

**7** **Going off-grid: Multiple benefits of rethinking the marking criteria format**

Tania Horák and Elena A M Gandini

---

**21** **Young learners' positive attitudes to learning in South America**

Amy Devine, Safiatú Lopes and Graeme Harrison

---

**29** **An investigation into English language teachers' experience of teaching adult groups online compared to face-to-face**

Siân Morgan, Carmen Ebner, Safi Lopes, Sally Moore and Anoushka Uzum

---

# Editorial

---

Welcome to *Research Notes 80*, which reflects on quality assurance in three different contexts of language testing. Both the assessment community and teachers involved in institutional assessment or preparing students for external exams will find useful insights and recommendations.

First, Horák and Gandini describe their small-scale trial of two marking formats for a high-stakes English for Academic Purposes (EAP) exam produced in their own higher education institution. They analysed how the contrast of marking by grid or checklist was perceived by raters, and the effect this had on rater reliability. The findings offer a strong basis for further refinement and implementation of revised marking methods.

Next, Devine, Lopes and Harrison discuss the concept of 'test ease', and present their study of young learners taking A2 and B1 exams in Argentina and Peru.

They explored levels of test ease, confidence, and motivation among candidates and found encouraging results. Implications for teaching practice, focusing on how to create the optimum learning environment, are included at the end of the article.

Finally, Morgan, Lopes, Ebner, Moore and Uzum share their findings from a study of English language teachers' perceptions and experiences of online lessons during the Covid-19 pandemic. Teachers reported they were able to achieve similar teaching aims to those of a face-to-face classroom, although these were sometimes achieved differently. The study demonstrates how existing assessment criteria for qualifications such as CELTA are fit for purpose for online teaching, but also the diversity of teaching experience and skill that must be considered as online teaching grows in use.

# Going off-grid: Multiple benefits of rethinking the marking criteria format

**Tania Horák** University of Central Lancashire, Preston  
**Elena A M Gandini** University of Central Lancashire, Preston

## Introduction

As per many other higher education institutions, the University of Central Lancashire (UCLan) produces their own set of English language exams, used, among other purposes, as part of the final assessment of pre-sessional English for Academic Purposes (EAP) courses. These exams are standardised and mapped against the Common European Framework of Reference for Languages (CEFR, Council of Europe 2001) and are used to confirm that students have reached a general level of proficiency suitable for starting the course of their choice. At Levels B2 and C1, the exams can be described as EAP-light, and draw on a broad range of topics not requiring any specialist knowledge to avoid favouring specific academic disciplines. In this paper, we discuss an exam development initiative and report on a small-scale research study in which members of the Exam Team recently investigated how two formats for marking criteria might affect raters' reliability, and their reaction to new formats.

## Background to the study and literature review

In this exam, receptive skills are tested through different types of selected response item and are relatively easy to mark using an answer key or optical mark recognition system. However, the grading of large numbers of scripts within a very short timeframe to allow the Admissions Department to process offers and allow students to join their destination course, is generally regarded as challenging. This is especially the case on summer programmes since the turnover of tutors is typically higher than usual and there tends to be a high proportion of new staff. It is also a period when levels of professional experience differ across the team. All these factors contribute to the lack of familiarity with institutional norms and requirements in the teaching team, and pose challenges to speedy and reliable marking of students' scripts. Since these are relatively high-stakes exams, threats to any form of reliability have to be a concern for the exams team.

Productive writing skills are assessed through a short semi-academic essay task (classic essay structure of introduction, body and conclusion but without referencing; see also Caplan and Jones 2019), currently marked using typical grid-style marking criteria, with axes of band/level (10 bands, in our case) and categories (three: content and appropriacy; organisation and cohesion; language accuracy and range). Analytic scales are deemed to be appropriate since, as stated widely in the literature (see, among others, Bacha 2001, Goulden 1994), they increase both inter- and intra-rater reliability by facilitating the rating process through 'highly structured guides that reduce personal choice' (Barkaoui 2007:87). While our criteria may be described as analytic, at present, one single grade is reported for the candidate's script, potentially encouraging a holistic approach to grading. The affordance of reflecting any spiky profiles (Hamp-Lyons 2016, Kings and Casey 2013, McLean, Perkins, Tout,

Brewer and Wyse 2012) of students' writing sub-skills through examination of the three core writing attributes categorised above is currently not realised.

In addition, comments made during standardisation sessions and other anecdotal evidence suggest that different raters might not be consistently applying equal weighting to the three categories, as intended. Instead it seems there is a tendency, most likely subconscious, to apply their own personal weightings to different aspects of the criteria according to their beliefs about the nature of academic writing, and given the single global score awarded, as stated above, there is no way of easily monitoring this.

While we could have simply changed practice to allocation of sub-scores for each category to address this, we were aware of other matters which needed addressing. For instance, the ongoing monitoring of how the current criteria were being used had highlighted a certain degree of ambiguity and inconsistency in the wording of some descriptors. In particular, we were concerned about the somewhat artificial creation of graded modifiers aiming to describe the difference between adjacent descriptors (for example, *excellent*, *very good*, *good*, *satisfactory*, etc.) or to the way certain features were not consistently handled across the different bands. For example, in our current criteria, spelling is only mentioned in Level C1 and B1, and is not mentioned in any of the bands of the B2 descriptors; likewise the style of text is considered only in terms of 'appropriateness to the nature of the task' for C1, then appears again only once where texts are expected to 'show sensitivity to academic writing style', and is not included again in the descriptors for the lower bands. This inconsistency is a common characteristic that has been highlighted as occurring also in the CEFR (as discussed, among others, by Alderson 2007, Hulstijn 2007, Weir 2005). This matter of potentially confusing wording is discussed during standardisation events, especially in the attempt to provide guidance to novice markers and those who, albeit experienced with assessment work, were less familiar with our criteria.

Despite standardisation, we noticed a tendency towards 'standardisation slippage'. Measures were taken to monitor and address this slippage in the form of regular benchmarked-script marking exercises during large-scale marking events such as at the end of pre-session courses. Nevertheless, we felt the problems with the criteria, above all the rather artificial grading of modifiers and their application, were very likely affecting scoring reliability, and thus were worth reviewing to see if we could enhance marking quality.

As well as boosting rater reliability, we had two further goals regarding washback and washforward (Coombe, Folse and Hubley 2007, Horák and Gandini 2019, North 1993). Currently the grids are rather unwieldy, composed of descriptors to cover the three CEFR levels we assess (demonstrating the overlap of the top and bottom of adjacent levels) and cannot be described as especially user-friendly. Greater clarity of how and where attributes of effective academic writing are represented in our criteria, and making this information more accessible through choice of wording and format to all users (e.g. markers, tutors, students), would contribute significantly to rating quality. The aim was for the revised criteria to become more of a pedagogical tool which could influence the EAP course syllabus (positive washback) and also assist students in continuing their academic language development after taking the exam by facilitating clear, well-structured feedback relating directly to their personal performance in relation to the marking criteria (washforward). In this article, we focus on the first goal of rater support to enhance rater reliability.

In 2017, we started looking critically and considered alternative formats for the grid, as the current format necessitated employing the language of unhelpful graded modifiers (frequency or gravity primarily) and potentially ambiguous wording simply for the sake of filling all (or a specified proportion of) the grid boxes. In addition, we wanted the criteria to provide sufficient guidance to markers about ensuring all aspects of the criteria were

considered equally and consistently. Our search for alternatives led us to re-examine the literature comparing various styles of marking criteria, and we were particularly drawn towards performance-driven approaches as a way to go 'off-grid'.

Examples of the application of such approaches can be seen in various studies, in particular Upshur and Turner (1995) and later Fulcher, Davidson and Kemp (2010). The latter produced a format known as a decision tree. This rating instrument consists of a series of questions regarding specific aspects of the performance that the rater responds to in a given order. The wording of questions should not leave much room for interpretation on the part of the rater but, rather, prompt a series of binary decisions. Points are awarded at each 'branch' which are then totalled to indicate overall performance.

Both Upshur and Turner's (1995) and Fulcher et al's (2010) methods, however, were produced in relation to assessment of spoken production skills. While there is clear overlap with assessing writing, both drawing on subjective judgement of language performance, the constructs being evaluated differ. Additionally, the contexts in which their work was applied also differed (general EFL and English for Specific Purposes (ESP) classes respectively) so with these studies acting as possible models for us, we adapted the concepts for our own purposes.

## Outline of the study

In order to produce new criteria, we started from an analysis of the extant marking criteria. We used colour coding to identify each criterion in its various iterations across the grid to facilitate the production of a comprehensive list. This helped to capture features which were not included consistently across levels, or even appeared to some extent within different main categories. These features were then reformulated into questions, ignoring the graded modifiers which appeared in relation to a single feature across different bands, as shown in Figure 1.

The questions were subsequently grouped according to the results of a previous research task undertaken in the very early stages of our criteria development project. In this exercise, a group of expert EAP tutors and exam writers were asked to reach a consensus on the most pertinent features of effective academic writing, and thus achieve a list of qualities we wished to see in students' work. Through brainstorming, discussion and distilling, without recourse to any other extant sets of marking criteria, the seven categories shown in Figure 2 ('new categories') were agreed on.

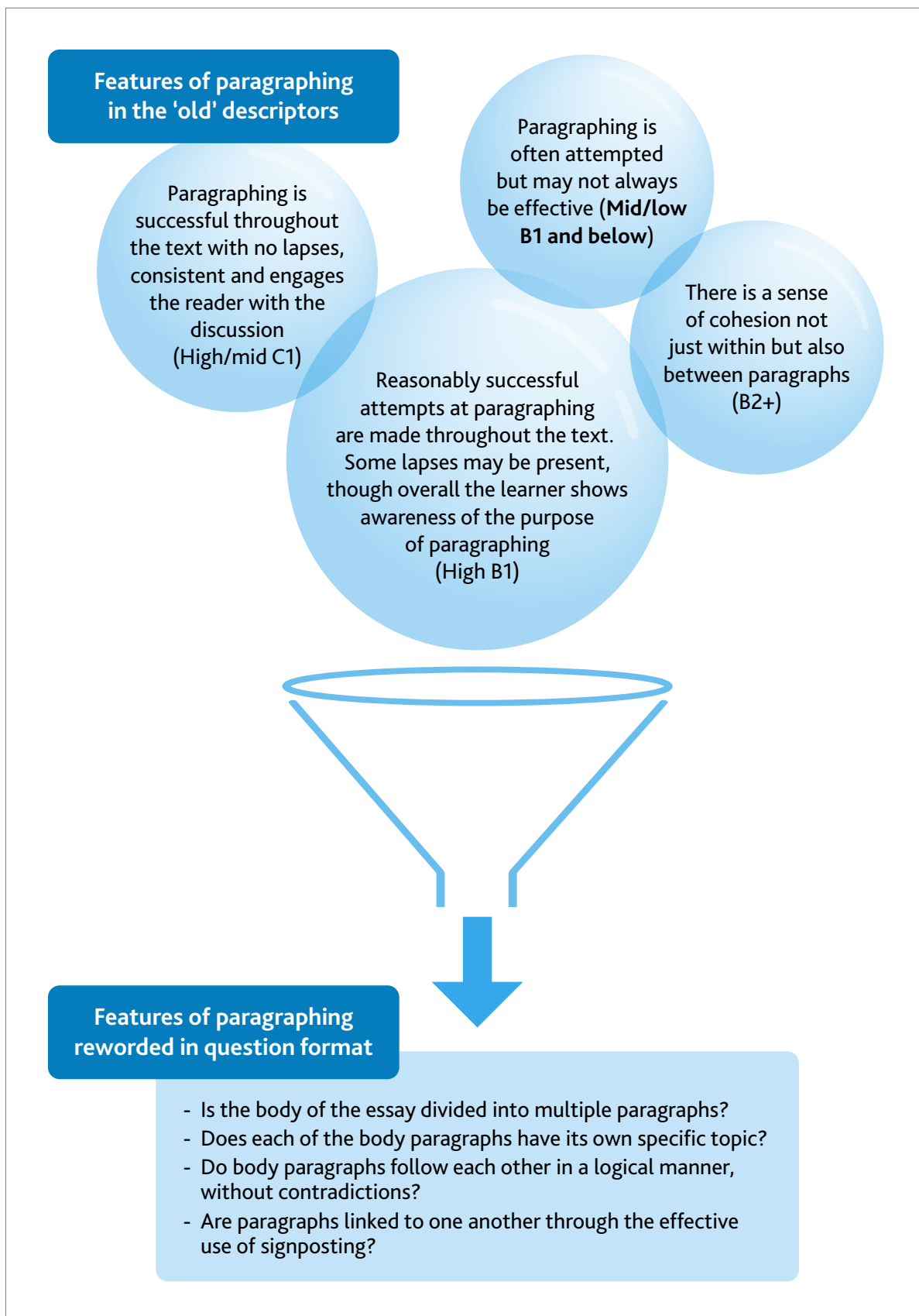


Figure 1: Example of the evolution from descriptors to question forms





**Figure 2: Main categories in the current grid-shaped writing criteria and in the new format (checklist)**

The first adaptations, shaped as decision trees as per Fulcher et al (2010), were not particularly favoured by colleagues, who found the branching hard to navigate. Therefore, the final outcome was shaped as a checklist rather than a tree, which would still allow for more guided marking and ensure no component would be overlooked. Furthermore, the checklist format had the additional bonus of potentially being more easily usable with students in class, thus contributing to our goal of facilitating improved feedback.

Our review of the literature had proposed possible solutions in the form of a guided approach afforded by a step-by-step rating instrument aimed at addressing the issue of idiosyncratic approaches to grade allocation, described previously. Consequently, we undertook a small-scale study comparing the two formats (grid v. checklist). One part of the study addresses the effect of format on scores and this will be reported elsewhere. To investigate the raters' perceptions of the formats, we formulated the following research questions (RQs).

RQ1: How would raters perceive the two types of criteria when asked to implement both?

- RQ2: Would raters perceive any advantages to checklist-style criteria?
- RQ3: Would the checklist-style criteria provide enhanced guidance for inexperienced raters?

## Methodology

The methodology we used for this very small-scale study consisted of requesting 37 participants to grade eight student essays. Four of these essays were to be graded using our former grid-style criteria and the other four were graded using our newly designed checklist-style criteria. The essays had been written by students on pre-sessional courses at UCLan; all responded to the same prompt. They had all previously been double-marked by members of the Exam Team to establish an agreed grade (from 0–10). Before commencing the grading of the eight essays, all participants were asked to review one particular essay together with the grade it had previously been awarded by the Exams Team in order to provide an indication of how performance mapped onto our criteria.

We recruited our participants through convenience sampling, tapping into our personal network of EAP tutors which comprised staff both at our own institution and elsewhere (see Table 2, described as internal and external contacts). We sent all the material via email and did not regulate whether our participants marked online or downloaded the essays to mark paper copies, in order to allow them to follow their usual marking practices.

In order to avoid order effect, the participants were divided into two groups. Group 1 were directed to use the grid first to assess a batch of four essays and then moved on to use the performance checklist for the second batch of four. Group 2 worked vice versa (checklist first, grid second). See Table 1. Thus, each of the eight essays was assessed using both types of criteria. There was a 76% response rate from our group of markers who had initially indicated willingness to take part (see Table 1).

**Table 1: The distribution of the essays amongst markers and returns**

	Grid	Checklist	Materials sent (N)	Responses received (N)
<b>Group 1</b>	Essays 1–4	Essays 5–8	18	16
<b>Group 2</b>	Essays 5–8	Essays 1–4	19	12
			<b>Total</b> 37	<b>Total</b> 28

Following their grading, participants completed a score sheet (in a simple Excel spreadsheet). The analysis of this quantitative data, including any possible effect of the imbalance between the two groups, is not reported here due to lack of space but we aim to report on this subsequently. Here we aim to focus on the perceptual data collected through a short questionnaire, completed after rating, in which our participants compared the implementation of the two forms of marking criteria (a copy is in the appendix). A series of open questions probed which style of criteria they preferred (following Barkaoui 2007) and also the perceived advantages and difficulties of each style. The questionnaire consisted of open questions to allow for personal explanation from each respondent. Some participant information was also collected for comparison purposes in the form of multiple-choice questions covering biodata and teaching and examining experience.

## Data analysis

All our participants were qualified EFL tutors but represented a range of experience in terms of teaching EAP, marking academic essays, working for major exam boards and also producing marking criteria themselves. From their questionnaire responses, participants were assigned to one of three groups, representing the relatively inexperienced, general EFL experienced and strong EAP experienced (Groups A, B and C) – see Table 2. This range of experience mimics many typical pre-sessional tutor group profiles. We were aware that level of experience may influence rater judgements of the rating methods presented given that it is known to affect rater performance (Davis 2016) and, thus, it was a benefit that we had representatives of each group.

The questionnaire responses were analysed using basic thematic analysis then negative case analysis (Mason 2002). Due to the relatively small amount of data, CAQDAS software was not applied to assist in the coding. Instead, simple colour coding in Excel was implemented. The themes which emerged were then grouped in light of how they addressed the three RQs, and key themes are reported below.

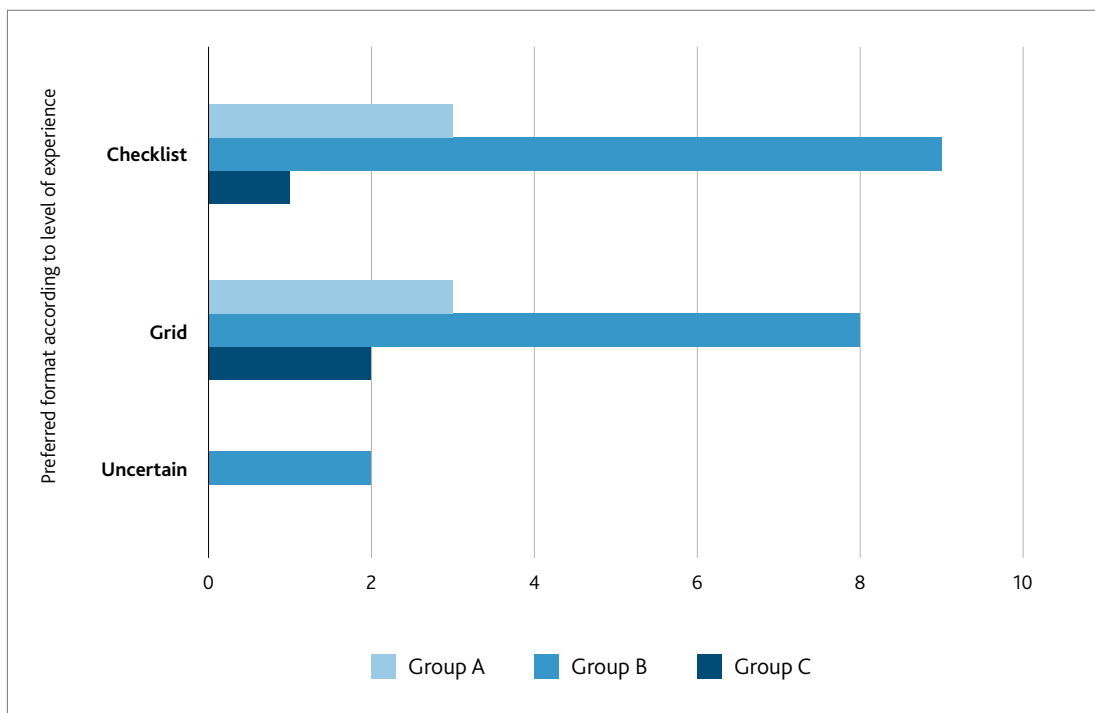
**Table 2: Tutor profile**

		N	%
Source of contacts	Internal contacts	n=10	35.71%
	External contacts	n=18	64.29%
Level of experience	Inexperienced (Group A)	n=6	21.43%
	General EFL experienced (Group B)	n=19	67.86%
	Strong EAP experienced (Group C)	n=3	10.71%

## Results

### RQ1

With RQ1 in mind (perceptions of the two formats) results revealed some rather frustrating outcomes. We had hoped for a clear preference either way. However, there was an even split (see Figure 3). Only a minority of the participants (7.14%) reported uncertainty regarding their preference. Participant 5, for example, stated benefits of each format but was unable to choose between them. Various Grid-Preferers (henceforth identified as, for example, PP1 G i.e. Participant #1 – Grid Preference) saw the grid as more holistic (PP4 G, PP14 G, PP18 G, PP19 G and PP25 G). On the other hand, nearly a third of the Checklist-Preferers (from now on referred to as PP2 C i.e. Participant #2 – Checklist Preference) commented that the checklist was more focused (PP11 C, PP16 C, PP17 C, PP27 C).



**Figure 3: Format preference of each experience group**

One theme addressing RQ1 (perceptions of each style) comprised comments regarding a single feature attributed to both styles. Firstly, **cognitive load** in implementing them was mentioned: members of both preference groups reported that there was too much information to juggle. For instance, PP18 G stated: 'Too many questions to consider and to keep in mind. For every essay I had to read them again' and PP4 G: 'So much information to digest!'. PP19 G made similar comments. Equally, PP17 C reported: 'Too wordy and tiring to read/remember'. PP16 C commented on similar lines. Each case was criticising their 'dispreferred' option. On the other hand, PP3 G said: 'Only 3 areas to consider so it's manageable.' and PP4 G commented on similar lines, on the grid. The checklist was found by some to be easier to manage: 'this style [checklist] requires less "cognitive" effort' (PP16 C) and '[it was] 'easier to focus on criteria in turn' (PP17 C). Raters who were familiar with grid-style indicated a preference based primarily on that familiarity (e.g. PP3 G, PP4 G, PP 14 G, PP 19 G, PP20 G), the first two being from the EAP-experienced group (C). Lack of familiarity and cognitive load are inevitably intertwined but only partially account for why it was variedly perceived.

Secondly, both groups made the claim that the method they preferred was **quicker**, as explained by PP8 C: 'I think it is easier and it doesn't take as much time to mark an essay as does the grid-style.' PP4 C, PP9 C and PP16 C made similar comments. Meanwhile, PP4 G, PP19 G, PP21 G, PP22 G and PP28 G in contrast, commented on the speed and/or the time saved by a grid, the latter also commenting that the checklist was for them quite time-consuming, echoed by PP14 G and PP20 G. Nevertheless, PP1 G acknowledged the possibilities of the checklist: 'It could be quick and easy once you got your head around [it]'.

Additionally, the level of **detail** the checklists provided was acknowledged in positive terms by PP2 C and PP13 C but even PP3 G, PP14 G, PP20 G, PP22 G (Grid-Preferers) saw the level of detail as an advantage. Only PP28 G felt there was too much detail in the checklist. While detail was discussed in relation to the grid, it did not appear to be as pertinent a theme (PP8 G, PP27 C, PP28 G).

## RQ2

RQ2 considered benefits of the checklist and in relation to this the **potential for feedback** to students was raised. PP22 G felt the checklist was 'better for feedback to students and supplying points for further improvements.' PP3 G commented likewise. Notably, they highlighted this advantage in the style they had not selected as their preference. Furthermore, PP7 C stated: '[it was] good for diagnostic info for students' and PP28 G concurred, reinforced by PP5 who added it could be 'a guide for improvement of discrete aspects of writing'. This point was important to us as the ability to produce more useful feedback contributes to our washforward goal. The mechanism by which improved feedback may be better delivered was suggested by PP4 G in that '[the checklist] might seem more transparent to students/stakeholders.'

There was evidence that our original intention of facilitating **consideration of all criteria equally** and consistently was achieved to some degree, which was encouraging, as PP19 G reported that the checklist 'reminds the marker of everything that we need to consider' (and PP25 G made similar remarks). PP4 G added to this: 'They [the questions] help you to remember aspects you might have forgotten about/don't feel are important'. Equally, PP21 G found 'it gives the rater a precise menu in what to look for in the answer/response'. Once more, it is notable they were all commenting on the style they had not chosen as their preference.

On a closely related note, **systematicity** appeared as a theme especially from the Checklist-Preferers (e.g. PP11 C, PP15 C, PP17 C); however, PP14 G also talked about the checklist helping markers think about the micro details of each section of the criteria and consider what contributes to each e.g. which elements exactly contribute to accuracy. Systematicity was not mentioned regarding the grid.

## RQ3

Addressing RQ3 (enhanced guidance), participants, unprompted, mentioned the **reduction of subjectivity** but different participants saw this feature in their own preferred style of criteria. For example, PP4 C commented: 'They [the questions] help you to remember aspects you might have forgotten about ... and so this reduces subjectivity'. Equally, PP13 C commented similarly that: 'I felt that using the check-list style marking criteria leaves less room for interpretation and subjectivity on the examiner's part' as did PP6 C: 'The questions led me so I might have been less biased'. From the other group, PP1 G said: 'I felt the grid was easier to use and less subjective'. What is more, participants from both preference groups acknowledged the benefit of reduced subjectivity in their dispreferred format; PP16 C stated: 'It [the grid] gives ... a more comprehensive overview of the criteria ... making marking less arbitrary', while PP18 G felt the checklist was 'possibly fairer and more objective'. Overall, therefore we can see that our participants saw the possibilities of reducing subjectivity via both formats; they were not seen exclusively as being offered by the checklist.

Other useful themes emerged regarding guidance (RQ3), not always positive. For example, it became clear that while a checklist format may have had advantages, some **wording** still needed interpretation, despite our attempts to address this. For instance, PP10, who did not make a style preference, said that a difficulty of the checklist was: 'interpreting the likes of "effectively", "sophisticated", "effective use". What is a "clear stance"? What is "undue repetition"?' PP10 also queried wordings utilised in the grid, however. Additionally, PP25 C found some sections had very similar wording and PP25 G likewise added: '[wording] made it hard to match students' varying skills to a grade.' Thus, although not widely mentioned, any lack of clarity was a concern.

## Rater experience

Having analysed the data in terms of the two preference groups, we investigated the themes by experience group but this did not appear to be a factor influencing perceptions. A few examples were notable, however. As already

noted, PP28 G found the amount of checklist detail unhelpful while also finding the grid insufficiently detailed. This might be explained by PP28 G's relative inexperience (a member of Group C) in using marking criteria and marking exam scripts. Also, two of the three most experienced participants were the ones who mentioned familiarity being influential in their grid preference. Being a very small group, it was not seen as a strong enough theme to be informative.

## Limitations

While we gained some valuable results, we recognise our study, like most research, suffered from certain limitations. First of all, we took the decision not to include a standardisation exercise with the participants. We would have had to standardise all participants with both sets of criteria and due to time (in terms of participants' workload burden) and logistical constraints, we felt this was not viable. However, all marking criteria need 'unpacking' and interpreting and several comments to this effect were made (PP3 G, PP4 G) so a form of standardisation will be put in place for the next round of piloting.

One of the main aims of our checklist was to reduce the amount of interpretation required of most marking criteria, and hence the need for thorough and ongoing standardisation measures to ensure a common (or as near as possible) application of the criteria. We were aware even during the production of the checklist that a certain level of interpretation was still required, however (e.g. 'Are the majority of tenses used accurately?'). With standardisation measures in future piloting, as mentioned, we hope to reduce the level of interpretation required.

As with all questionnaire data, we did not have the full picture behind the responses, and some would have benefitted from subsequent interviews to better interpret the comments (e.g. PP18 G: 'This is to me is a purposeful and pragmatic way of grading'). The limited scope of the study (this simply being an initial pilot) and the tight timeframe we were working within precluded this, however.

Finally, the outcomes of the ratings provided by our participants did not carry high stakes as they usually do when marking in a 'live' situation, so we were concerned the way the participants marked, influenced by lack of authenticity of the marking situation, may have affected the results unduly. This, however, was beyond the scope of our powers to control for and must simply be acknowledged.

## Discussion

Our study was small-scale but it fulfilled our aims in allowing us to evaluate our initial assumptions and ascertain whether the study had merit and our plan for altering the criteria format was worth pursuing. Overall, we felt the research produced useful outcomes since we gained a much better understanding of raters' interaction with grading criteria and how the format affected their decision-making. Despite two respondents (PP9 C and PP16 C) being enthusiastic about the checklist and claiming it posed no difficulties in using it, we must acknowledge a range of advantages had been voiced for both styles. Nevertheless, the support for our checklist from both inexperienced and experienced raters confirmed that this project was worth continuing. As a result, we felt we were nearer to achieving a workable set of criteria which would achieve our goals.

Secondly, to continue developing the checklist, it was clear we needed to further reduce the level of interpretation, for example, by introducing points per question, as in Fulcher et al's (2010) decision tree format, rather than points per category. All wording choices also need careful review. The perceived reduction of subjectivity afforded by the format was encouraging but needs further strengthening as it was judged to also be offered by the grid to some extent.

More importantly, change management relating to the implementation of any innovation is as important as the innovation itself with regard to successful adoption (Markee 1993). A change in practice incurs added cognitive effort and the emergent theme of familiarity accounting for grid preference alerts us to the peril of overlooking a common stumbling block for the implementation of new approaches. Transmission of accompanying explanation of, and rationale for, the new approach, together with appropriate training, will need careful consideration for this obstacle to be successfully overcome.

## Conclusion

Our investigation into the use and effectiveness of our marking criteria by altering the layout confirmed the checklist-style format had certain advantages. However, our investigation also identified other important considerations when making changes, such as lack of universal acceptance by the raters. The outcomes were very encouraging and we intend to undertake further research to investigate the practicality and effectiveness of a revised checklist, amended in light of our findings. If that pilot proves successful, we can explore in a longitudinal study whether and how washback and washforward, our other chief concerns, have been promoted by 'going off-grid'.

## References

- Alderson, J C (2007) The CEFR and the need for more research, *Modern Language Journal* 91 (4), 659–663.
- Bacha, N (2001) Writing evaluation: what can analytic versus holistic essay scoring tell us?, *System* 28, 371–383.
- Barkaoui, K (2007) Rating scale impact on EFL essay marking: A mixed method study, *Assessing Writing* 12 (2), 86–107.
- Caplan, N A and Johns, A M (Eds) (2019) *Changing Practices for the L2 Writing Classroom. Moving Beyond the Five-paragraph Essay*, Michigan: University of Michigan Press.
- Coombe, C A, Folse, K S and Hubley, N J (2007) *A Practical Guide to Assessing English Language Learners*, Michigan: University of Michigan Press.
- Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge: Cambridge University Press.
- Davis, L (2016) The influence of training and experience on rater performance in scoring spoken language, *Language Testing* 33 (1), 117–135.
- Fulcher, G, Davidson, F and Kemp, J (2010) Effective rating scale development for speaking tests: Performance decision trees, *Language Testing* 28 (1), 5–29.
- Goulden, N R (1994) Relationship of analytic and holistic methods to rater's scores for speeches, *The Journal of Research and Development in Education* 27, 73–82.
- Hamp-Lyons, L (2016) Farewell to holistic scoring. Part two: Why build a house with only one brick?, *Assessing Writing* 29, A1–A5.

- Horák, T and Gandini, E A M (2019) Improving feedback through computer-based language proficiency assessment, in Becerra, N, Biasini, R, Magedera-Hofhansl, H and Reimão, A (Eds) *Innovative Language Teaching and Learning at University: A Look at New Trends*, Voillans: Research Publishing, 95–103.
- Hulstijn, J H (2007) The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency, *The Modern Language Journal* 91, 663–667.
- Kings, P and Casey, H (2013) *ESOL Qualifications and Funding in 2014: Issues for Consideration*, London: Association of Colleges/NRDC.
- Mason, J (2002) *Qualitative Researching* (Second edition), London: Sage Publications.
- McLean, P, Perkins, K, Tout, D, Brewer, K and Wyse, L (2012) *Australian Core Skills Framework: 5 core skills, 5 levels of performance, 3 domains of communication*, available online: [research.acer.edu.au/cgi/viewcontent.cgi?article=1011&context=transitions\\_misc](https://research.acer.edu.au/cgi/viewcontent.cgi?article=1011&context=transitions_misc)
- Markee, N (1993) The diffusion of innovation in language teaching, *Annual Review of Applied Linguistics* 13, 229–243.
- North, B (1993) Transparency, coherence, and washback in language assessment, in Sajavaara, K, Lambert, R C, Takala, S and Morfit, C A (Eds) *National Foreign Language Planning: Practices and Prospects*, Jyväskylä: University of Jyväskylä, Institute for Educational Research, 157–193.
- Upshur, J and Turner, C (1995) Constructing rating scales for second language tests, *ELT Journal* 49 (1), 3–12.
- Weir, C J (2005) Limitations of the Common European Framework for developing comparable examinations and tests, *Language Testing* 22 (3), 281–300.



## Appendix

### Comparing the implementation of two forms of marking criteria – Participant Questionnaire



Your name:	
Date:	

Q.1	Which of the marking criteria did you find easier to use? A) Grid- style or b) Checklist-style?
Q. 2	What do you see as the advantages of using grid-style marking criteria (if any)?
Q. 3	What do you see as the advantages of using checklist-style marking criteria (if any)?
Q4.	Were there any aspects of the checklist-style criteria you found hard to use? If so, what?
Q5.	Were there any aspects of the grid-style criteria you found hard to use? If so, what?
Q.6	Please give us any further comments (if you have any) on your experience of using the two types of criteria in this research task.

<b>Please let us know about your background. If any questions are not entirely clear then add a brief note to explain.</b>	
Q7.	How many years' experience do you have teaching EFL?
	Less than 1 year
	1–2 years
	3–5 years
	6–10 years
	More than 10 years
Q8.	How many years' experience do you have teaching EAP (English for academic purposes)?
	None
	Less than 1 year
	1–2 years
	3–5 years
	6–10 years
	More than 10 years
Q9.	How much experience have you had marking academic/ semi-academic student essays?
	None at all
	Some experience
	I am highly experienced
	I have done a great deal of marking in the past but do not currently do any
Q10.	Have you ever worked for any exam boards to mark essays by candidates who are taking standardised international English exam (such as FCE, CAE*, IELTS)?
	Yes
	No
	If no, go to Q.12
Q11.	When was the last time you undertook such examining?
	This year
	Last year
	2–4 years ago
	More than 5 years ago
Q12.	Do you have experience of producing marking criteria for student writing yourself?
	Yes
	No
If you answered yes, please give a few details.	

\*Now known as A2 First, C1 Advanced

Many thanks for your co-operation, time and effort.

Tania and Elena.

# Young learners' positive attitudes to learning in South America

**Amy Devine** Research and Thought Leadership, Cambridge Assessment English

**Safiatú Lopes** Research and Thought Leadership, Cambridge Assessment English

**Graeme Harrison** Partnerships, Projects, Policy and Events, Cambridge Assessment English

## Introduction

The Pre A1 Starters, A1 Movers and A2 Flyers exams aim to be the first steps on the Cambridge English Qualifications learning journey, provide motivation and boost confidence. Previous impact research in Spain and Vietnam has indicated that the exams motivate learners and help them understand their progress (Ashton, Salamoura and Diaz 2012, Breeze and Roothoof 2013, Khalifa, Nguyen and Walker 2012). However, previous research did not specifically investigate the impact of Pre A1 Starters, A1 Movers and A2 Flyers on learners' preparedness for higher-level Cambridge English Qualifications. This article reports findings from a small-scale study in Argentina and Peru on the impact of having taken A1 Movers or A2 Flyers on young learners' emotions towards forthcoming A2 Key for Schools and B1 Preliminary for Schools examinations, and their beliefs and attitudes related to learning English.

Cambridge Assessment English (henceforth Cambridge English) exams are largely the preserve of private sector education in South America. However, with the comparatively high cost of the exams, many parents, and the private schools which serve them, have questioned the benefit of paying for the Pre A1 Starters, A1 Movers and A2 Flyers exams over and above the other, often burdensome, costs of schooling. They point out that the exams do not have any external currency, i.e., they are not accepted by higher education providers or employers as proof of English proficiency.

The Cambridge English rationale for the exams for young learners, however, is not predicated on the recognition of the exams, but rather their ability to motivate students, and provide both teachers and learners with a structure and aims for their learning. This structure has been likened to a 'ladder' of learning, on which students can securely climb from exam rung to rung until they reach the top. Teachers, parents and students can be sure that learning is appropriate for the level and optimal for improvement. As a result, beginning with Pre A1 Starters, A1 Movers and A2 Flyers exams can have a positive effect on achievement in later, higher-level and higher-stakes qualifications.

The study investigated the following research question: Among young first-time A2 Key for Schools and B1 Preliminary for Schools candidates, what is the impact of having previously taken Cambridge English Qualifications on young learners' test ease and their attitudes towards learning English?

The key variable of interest was learners' test ease. In this study, test ease was defined as learners' level of comfort and confidence regarding their forthcoming A2 Key for Schools or B1 Preliminary for Schools examination. Test ease can be considered the converse of test anxiety, a fear of testing situations, which is known to negatively affect test performance (Zeidner 1998). In addition to test ease, several other variables that have been found to influence learner behaviour and language learning outcomes were of interest, including speaking confidence, beliefs about intelligence and goal orientation (Dunlop, Devine and Gallacher 2017). Speaking confidence is the

degree to which learners feel confident speaking in a foreign language (also known as willingness to communicate; MacIntyre, Dörnyei, Clement and Noels 1998). Beliefs about intelligence denote the underlying beliefs that people hold about the ability to learn or aptitude (i.e. intelligence), which generally range from strong incremental beliefs, e.g., that intelligence is malleable and changeable, to strong fixed beliefs, e.g., that intelligence is innate and not changeable (Dweck, Chiu and Hong 1995). Goal orientation can be broadly defined as the social reasons why people engage in learning tasks, with two main reasons having been identified: to develop competence or to demonstrate competence (Dweck 1986). In addition, learners' interest in learning English and their engagement in English activities outside of lessons were also captured. A questionnaire covering the aforementioned variables was administered to learners who were due to sit A2 Key for Schools or B1 Preliminary for Schools for the first time. The participants included learners both with and without previous experience of the Cambridge English exams for young learners.

## Methodology

### Participants

132 11- to 15-year-old learners from six schools in Argentina and Peru participated in this study. All learners were due to sit the Cambridge English A2 Key for Schools or B1 Preliminary for Schools exams. Within this sample were learners that had previously taken the Cambridge English exams for young learners (A2 Flyers: 25 learners; A1 Movers: 37 learners) and learners that had not taken these exams (non-Cambridge English exams: 43 learners; no other exams: 27 learners).

### Questionnaire and focus groups

All learners completed a questionnaire, in Spanish, that included a mixture of demographic questions and Likert-type items adapted for use with young learners from established scales. Ten learners participated in two focus groups. Focus Group 1 included four students that had taken A1 Movers or A2 Flyers previously. Focus Group 2 involved six students that had never taken Cambridge English Qualifications.

The questionnaire asked learners about:

- test ease
- beliefs about intelligence (whether learners believed their English abilities were stable or could be improved – broken down into four components)
- goal orientation (the extent to which learners' goals focused on mastering English and on comparing their performance with their peers – broken down into three components)
- speaking confidence
- speaking fear.

Table 1 below provides more detail about the aforementioned measures included in the analysis.

**Table 1: Hypothesised constructs included in the questionnaire**

Measure (construct)	Description	Items
Test ease	To what degree learners felt calm about their forthcoming exam	6
Beliefs about intelligence – aptitude	To what degree learners believe they are good at English	3
Beliefs about intelligence – hard work	To what degree learners believe that their English can be improved through hard work	2
Beliefs about intelligence – fixed	To what degree learners believe their English ability cannot be changed	1
Beliefs about intelligence – incremental	To what degree learners believe they can get better at English	1
Goal orientation – mastery	To what degree learners wish to understand and master the English language	3
Goal orientation – performance compete	To what degree learners are motivated to do better than their classmates	3
Goal orientation – performance prove	To what degree learners need to prove their English ability to classmates	3
Speaking – confidence	To what degree learners feel confident about speaking English	1
Speaking – fear	To what degree learners feel speaking in English is scary	1

In addition, the questionnaire and focus group interviews asked learners about their:

- exam experience (whether they had taken Pre A1 Starters, A1 Movers, A2 Flyers or another exam previously or had no experience of English language exams)
- interest in English (learners' reasons for learning English)
- engagement in English activities (what English activities learners reported doing outside of class).

## Analysis

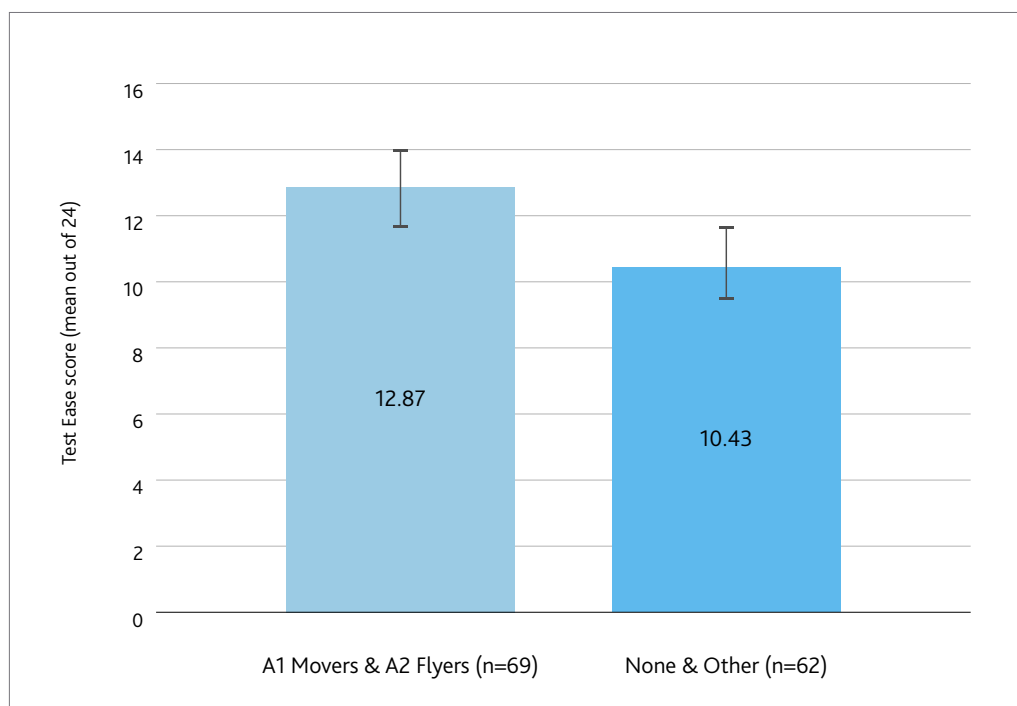
Data was analysed using SPSS Version 25. In general, test ease displayed a normal distribution of scores, whilst other constructs exhibited more skewed distributions. For this reason, non-parametric tests, such as Spearman's correlations, were applied. Regression analyses were also used to explore linear relationships between independent factors and test ease as an outcome. Focus group data was analysed for recurring themes.

## Results

The results suggest that learners that had taken A1 Movers or A2 Flyers<sup>1</sup> exams previously had higher test ease (lower anxiety) about their forthcoming exam than learners who had taken a different exam or who had never

1 No learners reported that Pre A1 Starters was their most recent English exam; however, some learners may have taken Pre A1 Starters in previous years.

taken an English language exam (see Figure 1). This effect was small, but statistically significant. Between these two exam experience groups, no further significant differences in beliefs and attitudes were observed.



**Figure 1: Mean test ease levels by exam group**

The majority of learners reported moderate to high levels of test ease (low levels of test anxiety). However, around 10% of learners reported high levels of test anxiety<sup>2</sup> and around 22% of learners reported a fear of speaking in English.

Learners reported positive attitudes towards learning English and the majority reported they were motivated to master the English language. Around 75% felt they were doing well in English and very few reported that learning English was difficult for them. Over 90% of students believed that they could improve their English language ability and that it could be improved with hard work. Only 30 to 40% of learners endorsed learning goals that focused on proving or comparing their performance to others. In addition, most learners reported regularly engaging in activities outside of English class to practise their English skills, although engagement levels varied by activity. The two most common activities reported by students were listening to music in English (87%) and watching TV and videos in English (80%). In addition, around half of the students regularly practised their English by surfing the internet (49%), talking to people (51%) and reading books and magazines (46%).

## Correlations

Correlations between all variables were examined, resulting in a total of 45 comparisons. A Bonferonni correction was applied to make the criterion for significance testing more stringent and decrease the chance of reporting a false positive relationship. Four correlations were found to be significant using this more stringent criterion (a p-value of .001 and below). These included:

---

2 Defined by the researchers as the learners with the lowest test ease scores on the scale in the learner questionnaire.

- 'Goal orientation – compete' and 'Goal orientation – prove' ( $r_s = .41$ )
- 'Beliefs about intelligence – hard work' and 'Beliefs about intelligence – mastery' ( $r_s = .37$ )
- 'Beliefs about intelligence – hard work' and 'Beliefs about intelligence – aptitude' ( $r_s = .39$ )
- 'Beliefs about intelligence – incremental' and 'Beliefs about intelligence – mastery' ( $r_s = .32$ ).

Although investigating the correlations between these variables was not a primary aim of the current study, the significant findings between learner beliefs and goal orientations are in line with previous research findings (e.g. Thompson and Musket 2005).

## Regression

Lastly, the degree to which each of the various learning beliefs and attitudes could predict test ease was explored through a hierarchical stepwise regression analysis. In a stepwise regression, all predictors are included in the model and SPSS then inspects which of the predictors are contributing the best to predicting the outcome in the model, thereby excluding predictors that add no significant value. Therefore, all variables (including exam experience group) were entered into the model at the same time, with test ease defined as the outcome.

The final model accounted for 20.8% of the variance in test ease scores (as indicated by the  $R^2$  value for the model). The final model only retained three variables, which were exam experience group, 'Speaking – confidence' and 'Goal orientation – prove'. This means that when all other factors were considered, these three factors combined offered the best level of prediction for the degree of test ease a student reported. In other words, having had experience with a Cambridge English exam (A1 Movers or A2 Flyers), and confidence in speaking in English, related to greater test ease. The latter may be indicative of the fact that the speaking component may be the greatest source of anxiety for some students, as previous research has suggested (Horwitz, Horwitz and Cope 1986). Additionally, having a strong goal orientation towards proving one's English performance had a *negative* relationship with test ease, meaning that those who reported being less interested in performing well in English in order to prove their ability level to other classmates also experienced greater test ease.

## Focus groups

The focus group results supported the quantitative findings – the young learners interviewed felt that learning English was very important; for example, learners commented (please note all comments are unedited to maintain authenticity):

*'I think English is one of the most universal languages and it is very important to know English.'*

*'For me English is increasingly important in too many aspects, whether economic, social, cultural, in any sense, is something basic, for example if you want to learn some work that has something to do with ... social entrepreneurship, you have to know English if ... or to communicate in an easier way ...'*

Many of the students indicated that they had intrinsic motivations for learning English, saying they liked it or were interested in the subject, but learning English was also linked to their future goals of obtaining jobs in, studying in, or traveling to English-speaking countries.

Similar to the questionnaire findings, the students described various ways in which they use English outside of their lessons. Example activities included listening to music with English lyrics, watching music videos, TV series

and films in English, communicating with other players whilst playing online video games, conducting online searches, communicating with friends and family members, translating for family members, and using English during travel to English-speaking countries.

Students reported varying degrees of anxiety related to their upcoming examination. Some described a lack of confidence or some nervousness about the exam as a whole or about the assessment of speaking or listening:

*'... I am a little bit anxious and nervous about the exam, I know it is not going to be easy at all ...'*

*'My problem is at the moment of speaking. Already speaking in Spanish or English I get nervous and I do not get the words. So, I'm nervous there already.'*

However, others expressed confidence about the upcoming examination and their comments suggested that preparation using practice tests contributed to their confidence:

*'I feel confident for the exam, I consider myself a person who can read the language easily in English, and I do not think I have many problems at the time of taking the exam.'*

*'... To be honest, I feel very confident. I have been progressing a lot with the practice tests.'*

*'... I feel like a mix of what everyone says, but we have been practising a lot with the practise tests and I have improved ...'*

Collectively, these results suggest that the young learners interviewed were generally not preoccupied by their forthcoming English tests.

## Limitations

Some limitations should be noted about the current research. Firstly, some constructs were measured with only one, two or three items in the learner questionnaire. Consequently, the internal consistency of some of the scales was low, suggesting that the items were not measuring the same construct or were measuring different aspects of the same construct. Unfortunately, we did not have the opportunity to pilot the questionnaire before conducting the research and the need to practically administer a short questionnaire measuring a range of affective variables was a key consideration in its development.

## Conclusions

The results of this study indicated that learners that had previously taken A1 Movers or A2 Flyers were calmer and more confident about their forthcoming English language exam than learners that had taken a different exam or no exam at all. However, the difference in test ease levels between the two exam experience groups was small. Having previously taken A1 Movers or A2 Flyers was also a significant predictor of learners' test ease, in combination with other variables. It is important to note that these findings should not be interpreted to mean that taking A1 Movers or A2 Flyers led to lower test anxiety in these learners, as there are other factors unmeasured in this study, such as the home and school learning environments, that may explain the association between exam experience



and young learners' emotional responses towards English language examinations. Nonetheless, the small but significant association between having taken A1 Movers or A2 Flyers previously and greater test ease is encouraging. Moreover, the majority of learners in this study held positive attitudes towards learning English, including motivation to master the English language and the belief that their English language abilities could be improved through hard work. Many learners engaged with activities outside of English class to practise their English language skills. Collectively these findings suggest that most young learners preparing for Cambridge English Qualifications in Argentina and Peru are highly motivated and engaged and generally have low anxiety about their forthcoming examinations.

## Implications for teaching practice

The learning environment plays a strong role in the development of beliefs about intelligence, learning goal orientations and attitudes towards assessment in young learners (Dunlop et al 2017). Thus, teachers and parents should aim to embody positive mindsets related to learning English, for example, by encouraging the belief that English (and other) abilities can be improved with hard work. They should also aim to create learning environments that inspire healthy goal orientations and learning strategies; for example, by encouraging learners to focus on learning goals, engage critically with feedback on their performance, and to develop effective help-seeking strategies. Teachers may be able to help reduce learners' anxiety about tests and fears related to communicating in English by encouraging a low-anxiety English language classroom (Young 1991). For example, teachers could:

- Enable learners to practice English in different contexts e.g., role-plays, games, in digital media, face-to-face with peers, with native speakers.
- Emphasise the importance of communication rather than perfect spoken accuracy; encourage young learners not to focus on their mistakes.
- Gradually encourage shy students to speak English in class (e.g., allow solo practice, then build up to small group work and speaking in front of the class).
- Avoid a high level of competition between peers in the classroom.
- Familiarise learners with the test format.
- Encourage learners to write down their worries or fears ahead of their tests (Ramirez and Beilock 2011) and to discuss any worries with their teacher.
- Introduce mindfulness training, which has been associated with improvement in students' resilience and stress (Zenner, Hernleben-Kurz and Walach 2014).
- Inspire the belief that English ability can be improved with hard work, and encourage healthy goal orientations (as suggested above), as doing so is likely to lessen learner anxiety as well.

## Acknowledgements

Special thanks to Maggie Dunlop, Bea Kalman and Andrea Canal for their contributions, and to Orange Day School, Oxford High School, Colegio Los Robles, Colegio San Patricio (campus in Sarandí, Avellaneda) in Argentina, and Colegio San Charbel and Colegio Alpamayo in Peru, for participating in this study.

## References

- Ashton, K, Salamoura, A and Diaz, E (2012) The BEDA impact project: A preliminary investigation of a bilingual programme in Spain, *Research Notes* 50, 34–42.
- Breeze, R and Roothoof, H (2013) *Final Report: Implantation of Cambridge English: Young Learners exams in Spanish schools: Teacher perspectives and washback on classroom practice*, Cambridge: Cambridge English Language Assessment.
- Dunlop, M, Devine, A and Gallacher, T (2017) *The role of learner characteristics in language learning: A literature review*, Cambridge: Cambridge Assessment English.
- Dweck, C S (1986) Motivational processes affecting learning, *American Psychologist* 41 (10), 1,040–1,048.
- Dweck, C S, Chiu, C and Hong, Y (1995) Implicit theories and their role in judgments and reactions: A world from two perspectives, *Psychological Inquiry* 6 (4), 267–285.
- Horwitz, E K, Horwitz, M B and Cope, J (1986) Foreign language classroom anxiety, *The Modern Language Journal* 70 (2), 125–132.
- Khalifa, H, Nguyen, T and Walker, C (2012) An investigation into the effect of intensive language provision and external assessment in primary education in Ho Chi Minh City, Vietnam, *Research Notes* 50, 34–42.
- MacIntyre, P D, Dörnyei, Z, Clement, R and Noels, K A (1998) Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation, *The Modern Language Journal* 82 (4), 545–562.
- Ramirez, G and Beilock, S L (2011) Writing about testing worries boosts exam performance in the classroom, *Science* 331, 211–213.
- Thompson, T and Musket, S (2005) Does priming for mastery goals improve the performance of students with an entity view of ability?, *The British Journal of Educational Psychology* 75 (3), 391–409.
- Young, D J (1991) Creating a low-anxiety classroom environment: What does language anxiety research suggest?, *The Modern Language Journal* 75 (4), 426–437.
- Zeidner, M (1998) *Test Anxiety: The State of the Art*, New York: Plenum Press.
- Zenner, C, Hernleben-Kurz, S and Walach, H (2014) Mindfulness-based interventions in schools – a systematic review and meta-analysis, *Frontiers in Psychology* 5, 1–20.

# An investigation into English language teachers' experience of teaching adult groups online compared to face-to-face

---

**Siân Morgan** Research and Thought Leadership, Cambridge Assessment English

**Carmen Ebner** Research and Thought Leadership, Cambridge Assessment English

**Safi Lopes** Research and Thought Leadership, Cambridge Assessment English

**Sally Moore** Teaching and Curriculum, Cambridge Assessment English

**Anoushka Uzum** Teaching and Curriculum, Cambridge Assessment English

## Introduction and context

In March 2020, the Covid pandemic made it necessary for educational institutions worldwide to transfer teaching online from one week to the next with minimal planning or preparation. This study, conducted by Cambridge English six months into the first lockdown, analyses English language teachers' experiences during this initial period, to find out whether teaching online had effects on classroom dynamics, lesson objectives, and learning activities, compared with teaching face-to-face (F2F). The inquiry focuses in particular on the teaching criteria of the Certificate in Teaching English to Speakers of Other Languages (CELTA). If significant differences between the two modes of delivery were found, the criteria would need to be revised in order to maintain the validity of the CELTA training course for an online teaching context.

A mixed methods research approach (Cresswell and Plano Clark 2011) was used to build a detailed picture of teachers' perceptions and experiences of teaching online. Data gathering took place in August 2020, by which time the initial crisis period had begun to settle and teachers had consolidated some online teaching practices.

Both quantitative and qualitative data were gathered from a survey publicised on a Cambridge English Teaching social media platform. The respondents (N=203) completed 28 questions to elicit their perceptions of whether the online medium enabled them to achieve similar teaching and learning aims to those of a face-to-face classroom. Respondents were also asked to report their levels of confidence in the effectiveness of teaching online, and identify any strengths or weaknesses it presents compared with F2F teaching. Follow-up focus groups were conducted remotely in October 2020 to probe more deeply into their experiences of teaching online.

## Brief literature review and research questions

Two main strands of literature were considered relevant to this study: computer-assisted language learning and digital literacy, and for the pedagogical aspect, studies from the language educational field, in particular learning-oriented assessment (Jones and Saville 2016, Turner and Purpura 2015).

Computer-assisted language learning (CALL) was defined as any process in which a learner uses a computer and as a result improves their language (Beatty 2003). Early CALL programs focused on learner autonomy, and used both CALL-specific software such as CD ROMs or web-based interactive software, as well as more generic software

applications such as Word and PowerPoint. Learning took place both synchronously and asynchronously, and was elicited through multiple-choice questions, quizzes, and matching, as well as simulations and web publishing. Significant early work in CALL was carried out by Warschauer (1996) and Warschauer and Healey (1998) and focused mainly on distance learning in higher education settings.

The literature on digital literacy is also relevant to this report (Hockly and Dudeney 2018, Prensky 2001). Prensky categorised users as either 'Digital Natives' and 'Digital Immigrants'; in his view, the latter would 'always retain, to some degree, their "accent" that is, their foot in the past'. Hockly and Dudeney instead discuss access to digital infrastructure, which they argue is more nuanced than simply a question of financial resources: 'language learning via mobile apps is a viable reality for many learners globally' (2018:166). They suggest that a digital divide is more likely to be linked to socio-economic factors, location, level of education, and age (Haight, Quan-Haase and Corbett 2014, Rideout, Foehr and Roberts 2010) and that such differences can exist within schools and even classrooms (Grant 2009, Warschauer 2004). Importantly, Yang and Egbert (2004) argue that definitions of the digital divide should also include the effectiveness with which technology is used as well as the skill level of the user.

Together with appropriate technological resources and digital literacy, second language pedagogy is a key consideration when F2F teaching is transferred to an online medium. Teaching has been described as 'a highly intricate socio-cognitive and sociocultural process' (Turner and Purpura 2015:261). It involves selecting and presenting appropriate input, elicitation, checking understanding, clarifying, monitoring, and giving both planned and unplanned feedback, all of which take place across a range of skills and interaction patterns. CELTA assessment criteria were developed to reflect this construct and, as stated, the aim of this study was to assess whether the online context enables CELTA novice teachers to provide evidence of these teaching skills and use of interaction patterns.

Since the 1980s, in large-scale English language teaching, learning and assessment has been underpinned by a communicative language model (Canale and Swain 1980), of which spoken interaction is an important component. Ensuring opportunities for natural interaction may be more effortful in an online context. For example, in the field of Conversation Analysis, turn construction and turn allocation (even in F2F lessons) has been described as 'a conundrum' for students (Wong and Waring 2010). This challenge could be greater in an online medium, where fewer visual cues may reduce opportunities for spontaneous interaction.

Using the principles of learning-oriented assessment (Jones and Saville 2016) with other learning technology may support aspects of F2F teaching which have proven difficult to transfer to online teaching. Digital tools can create opportunities for different types of interaction and different types of feedback and also reduce the assessment and administrative load on the teacher. These tools can also be used to store samples of performance for record keeping, and teachers can assess targeted aspects of performance both in real time and at their own convenience. Alternatively, automated marking tools can assist in self, peer, or group feedback, so that the learner can take an active role in the learning and assessment process.

The purpose of this study was to understand whether the CELTA criteria, originally developed for F2F teaching, were still fit for teaching online. Our research questions (RQs) were:

In an online context ...

1. Are teachers able to carry out with ease the same teaching activities as they do when teaching F2F?
2. Are teachers able to achieve the same teaching goals as effectively as they would do when teaching F2F?
3. Are there any teaching activities which are easier to do in an online medium than F2F?
4. Are there any teaching activities which are more difficult to do in an online medium than F2F?
5. Are the existing CELTA criteria fit for purpose to evaluate trainee teachers in an online environment?

## Methods and participants

A concurrent mixed methods approach was used (Cresswell and Plano Clark 2011). For the quantitative phase, a questionnaire was developed and constructs of interest operationalised during discussions between one member of Research and Thought Leadership (RTL) and two members of Teaching and Curriculum, departments in Cambridge Assessment English. The questionnaire contained a number of response types, including a 5-point Likert scale requiring a single response, and open-response items. The questionnaire was publicised via a Cambridge English Teaching social media platform, and administered through SurveyMonkey. A total of 203 teachers completed the questionnaire. The qualitative strand of the study consisted of the open responses throughout the questionnaire to enable respondents to identify themes not captured in the closed-response questions. Focus groups were conducted in October 2020 with four teacher/teacher trainers who had responded to an invitation inserted at the end of the questionnaire. Both the open-ended questionnaire responses and the focus groups were analysed using a qualitative approach of thematic analysis of the data (Braun and Clarke 2006).

## Survey findings

### **Participants: experience, professional development and teaching context**

Participants were a self-selecting group of teachers who had responded to an invitation on social media for teachers, so the first section of the survey gathered information on their experience, professional development and teaching contexts.

77% of the respondents reported over 10 years' and a further 13% reported more than five years' teaching experience. In terms of professional development, 55% respondents held a CELTA or Trinity CERT, both basic teaching qualifications, while 48% reported higher-level qualifications such as a DELTA (Diploma in Teaching English to Speakers of Other Languages) or Trinity diploma.

#### *Length of experience teaching online*

Over 56% of the respondents reported less than six months' experience teaching online and a further 20% between six months, and one year. It is likely that this reflects the emergency response to Covid-19, necessitating over half of the teachers in this sample teach online with little or no training.

### *Main teaching context/age group taught*

Respondents were asked to identify the context in which they taught most, and to respond to survey questions in relation to this context only. Most identified private language school (55%) and state education (32%). In terms of age range taught, 73% of respondents reported teaching adult learners, and 13% reported teaching students aged 15 to 18 years.

### *CEFR level taught*

In this survey, 63% of respondents reported teaching Levels B1/B2, with the remaining responses equally divided between A1/A2 (19%) and C1/C2 (19%). One potential future line of enquiry is whether certain CEFR levels can be taught more effectively online than others. However, in this study, the percentages for the A and C levels are too low to draw any meaningful conclusions about how these levels might be compatible with effective online teaching.

### *Class size*

Class size may influence the types of teaching activity teachers consider manageable when teaching online. The size of online classes reported ranges from below five to above 30, with most class sizes in this sample falling between five to nine and 16 to 20.

## **Teacher perceptions and experiences of teaching online**

### *Teacher confidence*

Our first question focused on teachers' levels of confidence in the overall effectiveness of online teaching. Of the 203 respondents, 42% reported feeling 'extremely or very confident' and 43% reported feeling 'somewhat confident'. These figures suggest that, despite the sudden and unexpected switch to online learning necessitated by Covid-19, most teachers were reasonably confident that online learning could be effective. However, a further 16% reported feeling 'not so confident' or 'not at all confident'.

### *Ideal number of students per class*

The study was aimed at teachers of adult groups, a typical teaching context for newly qualified CELTA teachers. In terms of class size, most responses fell between seven to eight and 11 to 12, relatively small numbers which may reflect the 55% of respondents who identified private language schools as their main teaching context.

The rest of the survey questions focused on the specific aspects of teaching as covered by the CELTA criteria. For reasons of space, a selection of representative responses is presented here. Please note all comments are unedited to maintain authenticity.

### *Planning online lessons and achieving teaching and learning objectives*

This matrix of questions probed teachers' experience planning online lessons compared with F2F lessons.

98% of respondents stated they were 'always or often' able to include similar learning objectives as they would for a F2F lesson.

84% of respondents stated they were 'always or often' able to include the same range of teaching objectives.

80% of respondents stated they were 'always or often' able to include the same range of learning activities.

Although responses were overall positive, respondents underlined the importance of detailed planning for online lessons, especially for trainee teachers:

*'You can be flexible but you have to have everything at your fingertips and anticipate what you might need.'*

In terms of pacing, 85% of respondents reported they were 'always or often' able to include changes of pace, similarly to a F2F lesson. However, some commented that the online environment made it more difficult to survey whole-class and group/pair activity simultaneously.

*'Pacing is much easier in a face-to-face course. You can see where the students are and how they're dealing with activities. This isn't easy online. When groups are in a breakout room, I can only enter 1 at a time and only get a glimpse at how the students are doing whereas in a face-to-face class I have a clear overview of what's going on.'*

In terms of planning time, only 44% stated planning online lessons 'always or often' took a similar amount of planning time as a F2F lesson. This theme also emerged in the open responses and focus group data, where respondents commented frequently that planning was more time-intensive.

*'It's all more time-consuming. Very much trial and error. Sometimes the process takes longer ... more steps involved. You have to be less ambitious. You have to prepare very carefully.'*

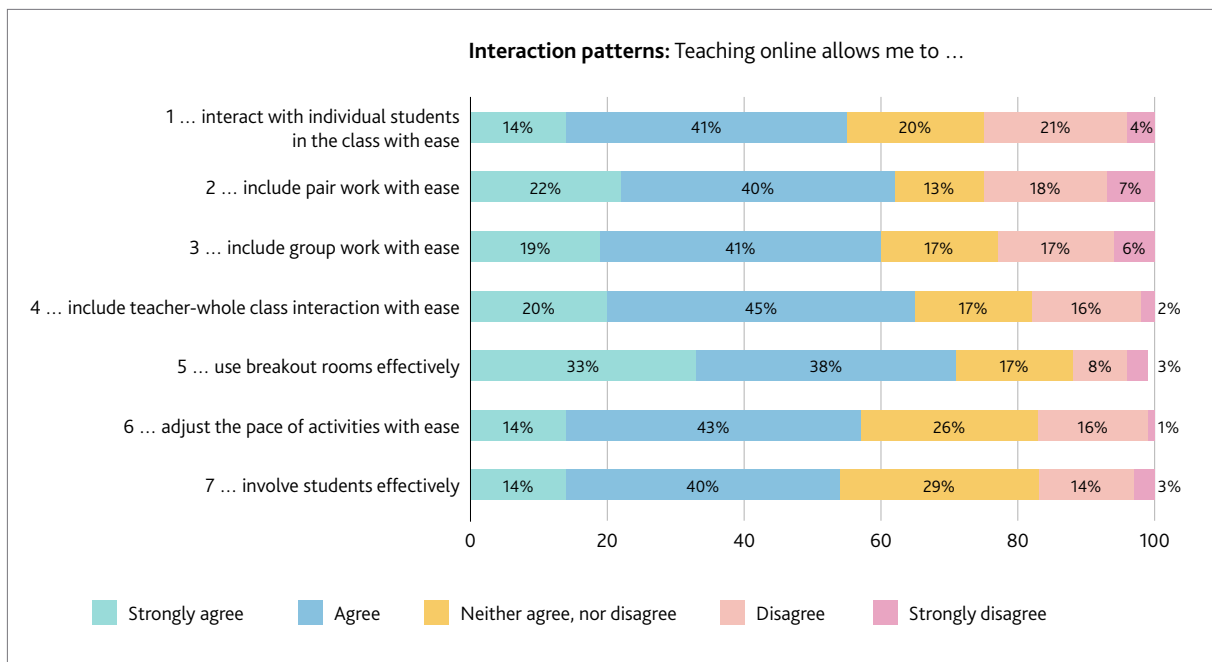
Some attributed this increase in planning time to the initial period of setting up, digitalising and resourcing F2F lesson material to online teaching, and anticipated it would decrease with time. Despite this extra time burden, however, there were several unsolicited comments on the professional satisfaction derived from the challenge of switching to online teaching.

*'I was not familiar with the virtual learning tools and not very tech savvy, thus it was a real stress and time consuming. However it was fun and worth the effort as I start falling in love with the remote learning and teaching.'*

*'It has been a most enriching experience with students responding well and brilliant attendance.'*

### *Range of interaction patterns used online*

Question 11 probed the type of interaction patterns teachers were able to include in online teaching, and the perceived ease and effectiveness with which they were used. As can be observed (Figure 1), all types of interaction formats scored moderately well, although the levels of agreement were slightly lower than for the previous section, and decreased as interaction moved from teacher-whole class to teacher-individual.



**Figure 1: Teachers' perceptions of interaction patterns**

The lower scores in interaction patterns may be due to a number of factors, both technical and psychological, raised in the open responses:

*'Group work is possible with breakout rooms but setting it up and providing access to the materials is slow.'*

*'Whole class interaction is not 'with ease' because of the need to mute/unmute. This kind of thing slows up the pace.'*

*'What does suffer is the kind of activities like "quickly compare answers". You can't keep putting them in and out of breakout rooms so that suffers a bit.'*

*'I think the activities are inevitably a bit more teacher centred.'*

These comments suggest that, for some teachers, including a range of interaction patterns in the classroom requires greater effort and forward planning.

### Teaching specific skills online

We wanted to probe teacher perceptions of how effectively they could teach online the different components of language ability: productive skills, receptive skills, functions, grammar and vocabulary, and pronunciation. Overall, there were good levels of agreement that online teaching allowed them to teach skills, functions, grammar and vocabulary, and pronunciation effectively.

#### Reading and writing

Over 72% agreed or strongly agreed that they were able to teach effectively all appropriate reading skills, and over 64% agreed or strongly agreed that they were able to teach effectively all appropriate writing skills. This was borne out in the interview data:



*'Reading goes a bit better. Students concentrate more and have a more intimate relationship with the text. What does suffer are the quick activities – "quickly compare answers".'*

*'With writing, if you set that up in Google docs in a breakout room, I've found that much more productive.'*

Some teachers felt that these could be effective if assigned for homework.

*'Students often feel they may as well do reading at home, especially higher levels from Latin languages.'*

*'Reading and writing seemed impossible to do during a lesson. We set that for work at home and corrected/had feedback during the lesson.'*

*'Writing is not encouraged online due to limited contact time. Use of flipped classroom approach means writing often set as part of pre-tasks.'*

### *Listening*

Using technology to teach listening skills was considered to be a natural fit by some:

*'You've got the chance to present something using some kind of board or presentation tool. Also you can use a digital version of the textbook which will incorporate the listening so that saves you having to do anything on the side using a recorder.'*

### *Speaking*

Over 76% of respondents agreed or strongly agreed that they were able to teach effectively all appropriate speaking skills. However, some commented that aspects of spoken communication could be compromised in an online medium by the quality of video/audio.

*'A large part of speaking and listening is about facial expressions and gestures and many of these subtleties are lost in an online classroom. For example, sound problems mean that participants are just trying to be heard as tone of voice and politeness go out of the window.'*

*'It is rather difficult when you're eliciting opinions ... you have to kind of just let them speak. I nominate them ... it is a bit tricky what comes out.'*

### *Grammar and vocabulary, and functions*

Over 70% agreed or strongly agreed that they were able to teach effectively all appropriate grammar and vocabulary and there were few comments regarding this aspect of online teaching.

Over 75% agreed or strongly agreed that they were able to teach effectively all appropriate functions online. However, some respondents expressed a need for more resources to do this.

## **Classroom management online**

### *Monitoring*

Many respondents reported positive experiences of monitoring students using small groups in breakout rooms.

*'Monitoring speaking can be more manageable while students are in pairs, or breakout rooms are easier. I believe that there's a lower level of anxiety for the student so they feel more comfortable in performing a task as compared to whole class monitoring. The teacher can also discuss the correction in more detail when students are in pairs or small groups.'*

Others gave examples of online tools they used to monitor skills work.

*'We've been using the annotation tool a lot more through the chatbox we found it quite useful. It allows them to do that anonymously so that nobody sees who's saying what.'*

*'You can use the chat function, to give feedback on emerging language speaking because I can ask students to interact at random and take notes of positive and negative features.'*

*'Monitoring writing is very easy when you have Google Docs as you can have everything open at the same time and have to deal with handwriting.'*

There were also examples of learning-oriented techniques used to monitor students' understanding in plenary mode.

*'I use the board to write a wrong answer then I ask my students what's wrong and we make some corrections together.'*

One concern was that the online medium narrows teachers' ability to visually monitor classrooms on several levels almost simultaneously.

*'It's hard to read body language as there are often students who can't be seen ... so there isn't the rich information you might get from the corner of your eye in a classroom (e.g. a pair chatting away when they've finished the activity).'*

*'Break out rooms can lead to uneven monitoring if the teacher gets involved in one BOR and doesn't get to the other.'*

Some respondents commented that the physical handling of the online medium reduced the opportunities for monitoring language:

*'It's harder to monitor for language use online as well – I'm not sure why, perhaps because there is a lot to processing to do at the same time (e.g. checking the chat, loading docs, sending links etc).'*

Other comments signalled a sense of audio and spatial disconnect between teacher and student which inhibited the traditional 'over the shoulder' monitoring teachers were used to in a physical online setting.

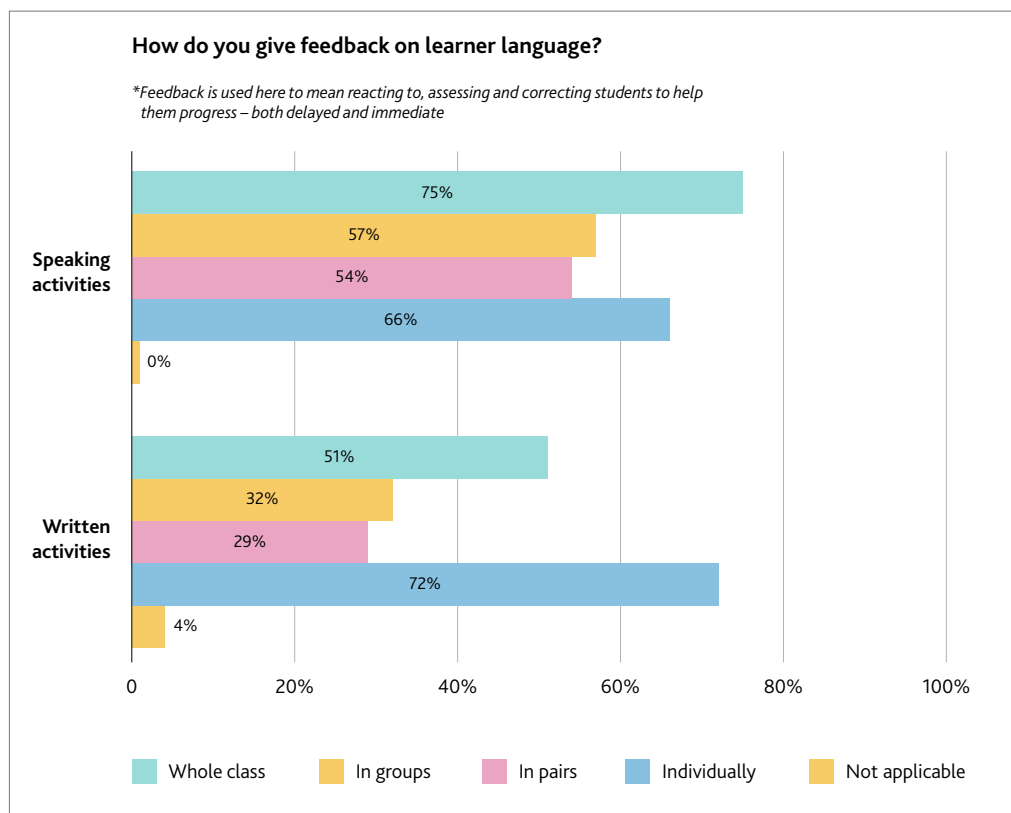
*'Monitoring anything online is more difficult due to a combination of the sound quality and the difficulty in knowing who said something. In a classroom, you have a sense of which direction a sound is coming from when monitoring groups or the whole class. In an online class, you have to be able to match the voice with the student.'*

## Feedback

Feedback is considered central to learning (Black and Wiliam 1998, Ramaprasad 1983). It can be provided during or after the learning activity, by the teacher, a peer, or the students themselves, and is intended to close the gap between current and future performance.

We were particularly interested in how and when teachers gave feedback for speaking and writing activities as these may require more individual analysis and attention.

Figure 2 suggests that respondents give feedback on speaking in a variety of interaction modes, with whole-class, individual, pair and group format all frequently used. The choice of format in giving feedback may depend on the task format (e.g. paired or group, task purpose, the timing of the feedback, whether it is planned or unplanned), as well as other contingencies.



**Figure 2: Teachers' perspectives on feedback**

There was more variation in format for giving feedback on writing, with the 1:1 format the most frequently used, but all formats were used to some extent. Here again, the feedback format selected may be dependent on a number of factors.

The open response data gave a rich picture of how teachers were adapting to the online context. Comments showed that feedback was indeed taking place with some creative use of technology and the immediacy it affords:

*'We can immediately show the products of the students on the screen and edit it together.'*

*'After the activity: usually in writing (chat or doc) – I type what I hear then ask students to self- or peer-correct.'*

*'Students send me their writing and I usually give them feedback by recording an audio.'*

### Timing of feedback

In terms of timing, feedback was given during and after the activity, as well as after the lesson. There was a preference in both speaking and writing skills for providing feedback after the activity, although 'in the moment' feedback did take place, particularly in speaking. Timing of and type of feedback seemed to vary according to a number of factors, as shown in the comments below:

*'It varies. I follow up speaking activities with some error correction as a whole class, perhaps using the whiteboard. With written activities, I give individual notes and then address the whole class with common errors in the next session. Sometimes I give immediate feedback during grammar or vocabulary work.'*

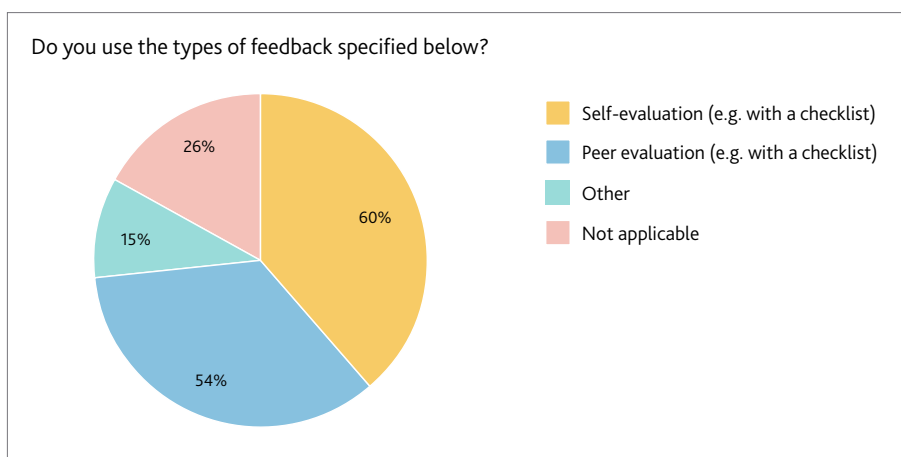
*'While explaining mistakes, I write the correction in the Google shared document (1 document is shared among us so as we can write simultaneously).'*

*'I start by giving them positive feedback and then focusing on a few mistakes. I write some sentences on the virtual board and have ss think of the mistakes and then in pairs find the right version'.*

*'Speaking: delayed feedback after fluency work. I share screen of what I have heard (good language) as well as mistakes and the students try to self-correct.'*

### Types of feedback

We also wanted to probe teachers' attitudes to learning-oriented techniques such as peer- and self-assessment as a feedback method. Positive attitudes to and experience with alternative forms of assessment would complement teacher feedback and could be integrated into online learning. This would enhance learner autonomy and lessen the burden of assessment on the teacher. Figure 3 shows a good proportion of respondents used self- and peer-evaluation with a checklist.



**Figure 3: Teachers' perspectives on types of feedback**

Respondents provided illustrative examples of learning-oriented activities in the open response data:

*'I provide samples of language mistakes for group to correct.'*

*'Group checklist of goals – presented at start of class and at end of class we give thumbs up to each goal we achieve.'*

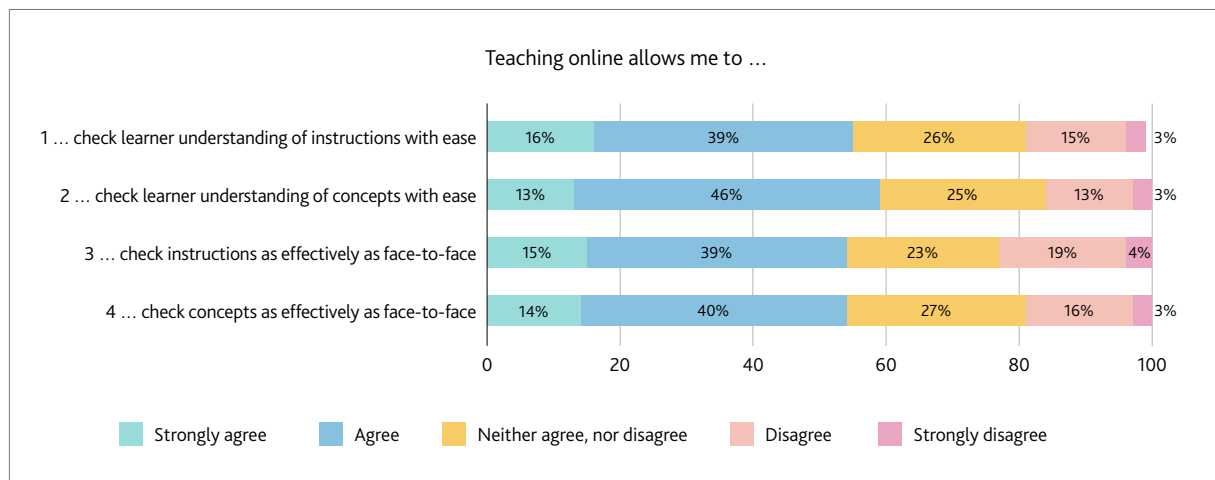
*'individual feedback or class feedback with a check list and with examples.'*

When asked whether they perceived giving feedback online to be as effective and time efficient as giving feedback F2F, over 60% of the respondents agreed or strongly agreed that giving feedback when teaching online takes more time than giving feedback F2F. However, 50% of the correspondents agreed or strongly agreed that giving feedback when teaching online is as effective for them as when giving feedback F2F.

These percentages suggest that, particularly in the initial phase of teaching online, teachers might benefit from support in the form of training and effective technological tools to reduce the time burden of giving personalised feedback. Another potentially useful form of support would be online communities (Lave and Wenger 1991) of practice set up for teachers to share their resources and expertise in a 'virtual staffroom'.

## Checking understanding

Embedded into classroom interaction and learning is the teacher's ability to check learners' understanding of both instructions and concepts. The following questions were designed to elicit respondents' perceptions of the ease and effectiveness with which they were able to do this online.



**Figure 4: Teachers' perspectives on checking learners' understanding online**

Responses generally indicated reasonable agreement with the statements, but also revealed levels of disagreement which indicate that checking understanding is another aspect of online teaching in which teachers would benefit from training and support.

## Elicitation language and concepts online

Eliciting is carried out by teachers to extract concepts and language from the learner, rather than providing the information themselves. As such, it is a key technique in checking and promoting learning. We were interested in finding out the ease and effectiveness with which teachers were able to do this in an online environment.

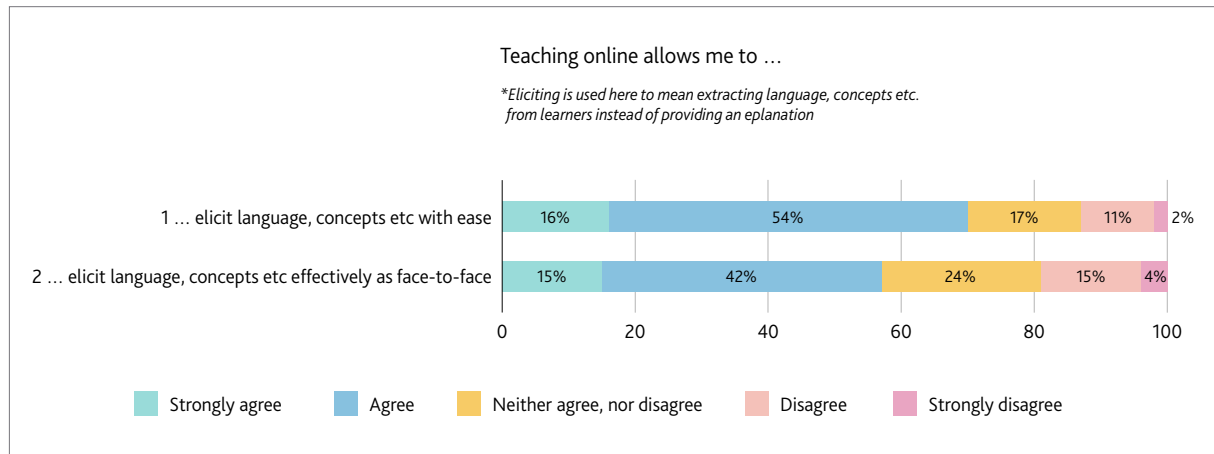


Figure 5: Teachers' perspectives on eliciting online

Figure 5 shows that 70% of teachers agreed or strongly agreed that they were able to elicit language and concepts with ease when teaching online. However, 57% of teachers agreed or strongly agreed that they were able to elicit language and concepts as effectively F2F as when teaching online, which suggests that checking concepts is another aspect of online teaching where teachers would benefit from training and support.

## Engaging with students online

One perception of online teaching respondents expressed concern about is that the remoteness of the medium puts a distance between them and their students which makes it difficult to engage and build relationships with them. Figure 6 confirms this concern.

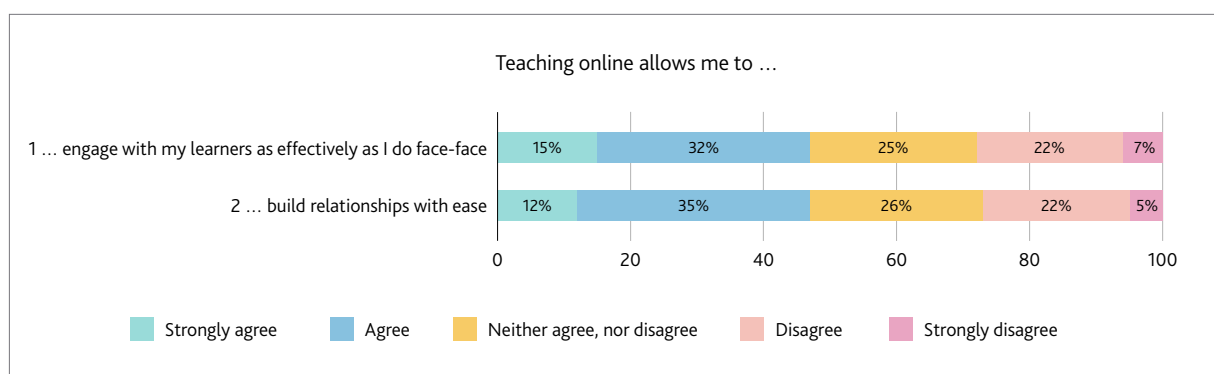


Figure 6: Teachers' perspectives on engaging with learners online

This is an area which may merit further training or resourcing.

## Perceived advantages and disadvantages of teaching online and face-to-face

Two final open questions were included to gather respondents' perceptions of the perceived advantages or disadvantages of teaching online, and teaching F2F.

### Is there anything that you do online that you would not be able to do in the classroom? If so, what?

Of the 203 respondents, 161, answered this question. Thirty stated that there were no advantages. Of the remaining 131, one major theme which emerged (37 responses) was the greater opportunity to use technology as a flexible resource, both as part of planned lessons as well as unplanned moments of contingency, e.g., to check a reference or give an example.

*'Lots of digital tools can be used during the class that students might not have access to in their physical classrooms. Similarly, they also have a host of realia at their fingertips and can show and tell more about their lives through physical objects from their home/wherever they are.'*

*'Collaborative activities work exceptionally well using digital interfaces and can be shared with other learners really easily and effectively. Flipped learning also works more seamlessly.'*

*'It's less of a challenge to have students work collaboratively using Google Docs/Padlet/a webquest etc. because the technology doesn't have to be brought into the classroom – it's already there.'*

*'As students are working online from home on their own computer/tablet they can complete online worksheets, games and exercises and get feedback. In the classroom I do not have enough laptops for everyone to do this at the same time.'*

*'Real time correction though Google docs.'*

As indicated in one of the earlier survey responses, there was considerable appreciation of the benefits of breakout rooms, which some respondents reported as enabling more personal interaction, and as less threatening for shy students.

*'It is much easier to hear learners when they are speaking in BO rooms and notice language.'*

*'Breakout rooms give privacy, allow own pace of work.'*

Other comments regarded the possibility of storing language samples: 'I can record the lessons – or parts of them – for learners to watch and monitor their own language use'; and the potential of online learning for teaching writing skills: 'collaborative writing that's genuinely collaborative.'

Despite the lower levels of agreement about the potential for engagement when teaching remotely, some comments indicated the contrary.

*'I connect with my learners at a much deeper level. Learners can share their personal space and parts of their lives (their pets, their family members, etc.). So, much stronger learning community.'*

*'Much more interesting discussions/conversations.'*

Several respondents expressed the opinion that teaching online is more restricted than teaching F2F, so the final questions were aimed at gathering examples of the kinds of teaching activities that teachers found more difficult to do online.

### **Is there anything that you do in the classroom that you can't do online?**

Of a total of 178 responses, 13 respondents answered 'no' or 'not really' to this question and two skipped it. Of the remaining 163 open responses, several themes emerged.

#### *Kinaesthetic activities*

One recurring example was that the online environment made it difficult for teachers to introduce kinaesthetic activities such as mingles, running dictations, or board games into their lessons (14 responses).

*'Move about, keep the energy levels up, respond to emotions students may be feeling that day, get involved in TPR/ kinaesthetic activities, draw students together, read actual books and papers, projects ...'*

*'The more kinesthetic activities (matching slips of paper for example) are difficult.'*

#### *Lack of physical/visual connection*

Another highlighted issue in online teaching was the lack of physical connection and eye contact with students. Some correspondents reported that this lack of visual cues could make it difficult to read the classroom, and monitor learner engagement or understanding (13 responses).

*'When I can't see students it's more complex to understand if they are not participating (tech issues, logged on and left), or not understanding.'*

#### *Speed and flexibility in responding spontaneously*

Also noted was a reduction in the speed and efficacy of executing classroom decisions in monitoring, managing interaction, and taking spontaneous classroom decisions (16 responses).

*'Pairing students with ease and frequently. I can pair students for a short project, but when I want students to "work with a partner" on an exercise, for example, I can't easily accomplish that in the online class.'*

*'React immediately, grab some extra examples from a book I know, quickly give examples on the board.'*

#### *Producing unplanned visuals*

Some respondents missed the opportunity to produce impromptu boardwork to react to contingencies.

*'I use the whiteboard a lot in my classes and I find that I can do a lot more on the large class boards than I can on a small electronic one.'*

#### *Interaction and engagement*

Finally, several respondents commented on the greater effort in developing and maintaining natural and low-key interaction, as well as building rapport and a class community.



*'Walk around the room, make eye-contact, give non-verbal praise/encouragement/feedback, see who is not participating fully and why.'*

*'Work in small groups, establish rapport.'*

*'Elicit, build a class community.'*

*'Use gesture and expression more effectively, keep a low profile more often, engage with students more naturally.'*

Respondents' attitudes to teaching online were operationalised in four measures: comfort teaching online, confidence in their own ability to teach English online, confidence in student learning online, and the range of digital tools used.

Overall, respondents reported positive attitudes to teaching online with more than 74% agreeing or strongly agreeing with the following statements:

- They felt comfortable with teaching online.
- They were confident that students learn when they taught online English.
- They were confident of their ability to teach online.
- They used a range of digital tools when they taught online.

Despite the suddenness of the switch to online teaching in mid-winter 2020 for teachers with mixed levels of online teaching experience and resourcing, the majority of respondents reported, six months later, high levels of confidence in their ability to perform most of their normal teaching activities through the new medium. Indeed, some saw the experience as transformational.

## Discussion

There are limitations to this study. Participants were a group of 203 self-selecting teachers who responded to an online survey in a social media platform. Of these, 20 randomly selected participants were invited to take part in the focus group, but only four were able to attend. Nevertheless, the resulting data, combined with the open survey responses, have provided rich qualitative accounts of teacher perceptions and experiences.

In terms of teaching profile, this was a reasonably homogenous group of experienced teachers of groups of adults, teaching mostly at B1 and B2 level. Many of these reported relatively little experience of teaching online, and we do not have data on their levels of general digital literacy, which, if low, may have influenced their perceptions of teaching online. However, there will be many more teachers in different contexts whose perceptions and experiences this study has not been able to capture; teachers who are comfortable teaching online, and others for whom challenges exist. Another possible limitation is that we did not collect data on learners' experience of being taught online, nor measure language learning gains during online teaching. This was beyond the scope of this preliminary study, and perhaps inappropriate during the initial unsettled period of lockdown teaching.

Nevertheless, findings from this preliminary study have identified common themes and provided information on both the challenging aspects of teaching online, as well as its unexpected positive impacts. Overall, respondents stated that they had been able to:

- cover curriculum aims
- cover all skills
- use a variety of feedback options
- use a range of interaction patterns
- use a range of digital options.

Furthermore, teachers commented favourably on the positive aspects of teaching online. For example, breakout rooms allowed students to work at their own pace, and that the privacy of smaller groups and absence of background noise lowered anxiety levels and increased confidence for some students. Some respondents commented on the ease of collecting and storing language samples, as evidence of student performance, or as a basis for student reflection, or to be used for self- and peer-feedback based on principles of learning-oriented assessment. Finally, many appreciated having online resources immediately to hand for impromptu activities; as one teacher put, it 'a host of realia'. In that sense, another respondent commented, online teaching reflects authentic real-life 21st century skills.

In terms of more challenging aspects of teaching online, a number of respondents reported:

- increased planning time
- greater difficulty in monitoring and giving feedback
- more teacher-centred lessons
- less dynamic interaction
- fewer opportunities to respond naturally to contingencies.

Despite these perceived disadvantages for some, many respondents stated they were confident of their ability to cover their teaching and learning objectives through the online medium.

## Preliminary conclusions

This study was originally carried out to investigate whether the original assessment criteria for CELTA were fit for purpose for assessing candidates teaching online. It offers a glimpse of the multitudinous concerns, competing priorities, teaching and planning decisions, and potential contingencies that teachers take into consideration before and during the course of every lesson, and which they have been transferring online during the 2020 Covid-19 crisis.

The data collected indicate that this cohort of teachers believed they were able to carry out all of the teaching activities and goals set out in the CELTA assessment criteria, although in some cases they achieved their aims differently. Therefore, the results suggest that the existing CELTA criteria, developed for F2F teaching, remain fit for purpose for the assessment of online teaching, as long as a gloss is provided to assist the examiners in the new context.

While the data gives a positive picture of online teaching overall, it also sends a clear message regarding the diversity of teachers' experience of certain teaching techniques and ease using online tools. The findings presented could be used to inform future support and training provision for the growing population of teachers adapting to this new paradigm, both in pre- and in -service training, and for ongoing professional development.

## References

- Beatty, K (2003) *Teaching and Researching Computer-assisted Language Learning*, New York: Longman.
- Black, P and Wiliam, D (1998) Assessment and language learning, *Assessment in Education* 5 (1), 7–72.
- Braun, V and Clarke, V (1996) Using thematic analysis in psychology, *Qualitative Research in Psychology* 3 (2), 77–101.
- Canale M and Swain, M (1980) Theoretical bases of communicative approaches to second language teaching and testing, *Applied Linguistics* 1 (1), 1–47.
- Cresswell, J W and Plano Clark, V L (2011) *Designing and Conducting Mixed Methods Research* (Second edition), California: Sage Publications.
- Grant, L (2009) *Learning in Families: a review of research evidence and the current landscape of Learning in Families with digital technologies*, available online: [www.nfer.ac.uk/publications/FUTL30/FUTL30.pdf](http://www.nfer.ac.uk/publications/FUTL30/FUTL30.pdf)
- Haight M, Quan-Haase A and Corbett B A (2014) Revisiting the digital divide in Canada: The impact of demographic factors on access to the internet, level of online activity, and social networking site usage, *Information, Communication & Society* 17 (4), 503–519.
- Hockly N and Dudeney, G (2018) Current and future digital trends in ELT, *RELC Journal* 49 (2), 164–178.
- Jones, N and Saville, N (2016) *Learning-oriented Assessment: A Systemic Approach*, Studies in Language Testing volume 45, Cambridge: UCLES/Cambridge University Press.
- Lave, J and Wenger, E (1991) *Situated Learning: Legitimate Peripheral Participation*, Learning in Doing: Social, Cognitive and Computational Perspectives, Cambridge: Cambridge University Press.
- Prensky, M (2001) Digital natives, digital immigrants: Part 1, *On the Horizon* 9 (5), 1–6.
- Ramaprasad, A (1983) On the definition of feedback, *Behavioural Science* 28 (1), 4–13.
- Rideout, V, Foehr, U and Roberts, D (2010) *Generation M2: Media in the Lives of 8- to 18-year-olds*, Kaiser Family Foundation Study, available online: [www.kff.org/entmedia/8010.cfm](http://www.kff.org/entmedia/8010.cfm)
- Sepulveda-Escobar, P and Morrison A (2020) Online teaching placement during the COVID-19 pandemic in Chile: challenges and opportunities, *European Journal of Teacher Education* 43 (4), 587–607.
- Selivan, L (2020) *3 ideas for synchronous online lessons on Zoom*, available online: [leoxicon.blogspot.com/2020/03/3-ideas-for-synchronous-online-lessons.html](http://leoxicon.blogspot.com/2020/03/3-ideas-for-synchronous-online-lessons.html)
- Turner, C E and Purpura, J E (2015) Learning-oriented assessment in the classroom, in Tsagari, D and Bannerjee, J (Eds) *Handbook in Second Language Assessment*, Boston: DeGruyter Mouton, 255–274.
- Warschauer, M (1996) Computer-assisted language learning: An introduction, in Fotos, S (Ed) *Multimedia Language Teaching*, Tokyo: Logos International, 3–20.
- Warschauer, M (2004) Technological change and the future of CALL, in Fotos, S and Crown, C (Eds) *New Perceptions for CALL for Second and Foreign Language Classrooms*, Mahwah: Lawrence Erlbaum Associates, 15–25.
- Warschauer, M and Healey, D (1998) Computers and language learning: An overview, *Language Teaching* 31 (2), 57–71.
- Wong, J and Waring, H Z (2010) *Conversation Analysis and Second Language Pedagogy: A Guide for ESL/EFL Teachers*, New York: Routledge.
- Yang Y and Egbert J (2004) Mediating the digital divide in CALL classrooms: promoting effective language tasks in limited technology contexts, *ReCALL Journal* 16 (2), 280–291.

# We help people learn English and prove their skills to the world

Discover more:  
[cambridgeenglish.org/researchnotes](https://cambridgeenglish.org/researchnotes)

We are Cambridge Assessment English. Part of the University of Cambridge, we help millions of people learn English and prove their skills to the world.

For us, learning English is more than just exams and grades. It's about having the confidence to communicate and access a lifetime of enriching experiences and opportunities.

With the right support, learning a language is an exhilarating journey. We're with you every step of the way.

Cambridge Assessment English  
The Triangle Building  
Shaftesbury Road  
Cambridge  
CB2 8EA  
United Kingdom

 [cambridgeenglish.org](https://cambridgeenglish.org)

 [/cambridgeenglish](https://www.facebook.com/cambridgeenglish)

 [/cambridgeenglishtv](https://www.youtube.com/cambridgeenglishtv)

 [/cambridgeeng](https://twitter.com/cambridgeeng)

 [/cambridgeenglish](https://www.instagram.com/cambridgeenglish)

 [/cambridge-assessment-english](https://www.linkedin.com/company/cambridge-assessment-english)